

# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 98891

TO: Maher Haddad  
Location: 9d08 / 9e12  
Wednesday, July 16, 2003  
Art Unit: 1644  
Phone: 306-3472  
Serial Number: 09 / 647544

From: Jan Delaval  
Location: Biotech-Chem Library  
CM1-1E07  
Phone: 308-4498

jan.delaval@uspto.gov

### Search Notes

- ① Page # 9, Line 27, Need seq ID No.  
② Page 15, check Amed By  
18 4 4 65  
3 30, Table I Need seq 4

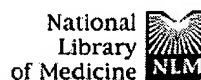
New matter  
seq ID No.  
11  
==

Examiner's notes

Jan Delaval  
Reference Librarian  
Biotechnology & Chemical Library  
CM1 1E07 -- 703-308-4498  
jan.delaval@uspto.gov

**THIS PAGE BLANK (USPTO)**





PubMed	Nucleotide	Protein	Genome	Structure	PMC	Taxonomy	OMIM	Books	
Search	PubMed	for						Go	Clear
Limits		Preview/Index		History		Clipboard		Details	

Display	Abstract	Show: 20	Sort	Send to	Text
---------	----------	----------	------	---------	------

1: Cytogenet Cell Genet. 1999;87(3-4):238-44.

[Related Articles, Links](#)

Entrez  
PubMed



## The integrin alpha10 subunit: expression pattern, partial gene structure, and chromosomal localization.

Lehnert K, Ni J, Leung E, Gough S, Morris CM, Liu D, Wang SX, Langley R, Krissansen GW.

PubMed  
Services

Department of Molecular Medicine, Faculty of Medicine and Health Science, The University of Auckland, Auckland, New Zealand.

Herein we report the cloning of cDNAs and incompletely processed hnRNAs from endothelia and heart that encode the alpha10 subunit forming part of the novel collagen type II-binding integrin alpha10beta1 of chondrocytes. Analysis of hnRNA clones and reported expressed sequence tags revealed the positions of 17 putative intron-exon splice junctions shared with those of the p150,95 (ITGAX) gene. Human alpha10 transcripts of 5.4 and 1.8 kb were not restricted to chondrocytes but, instead, were widely expressed in a panel of 24 tissue types, where the highest expression was found in muscle and heart. The human alpha10 subunit gene (ITGA10) was localized to band q21 of chromosome 1. Copyright 2000 S. Karger AG, Basel

Related  
Resources

PMID: 10702680 [PubMed - indexed for MEDLINE]

Display	Abstract	Show: 20	Sort	Send to	Text
---------	----------	----------	------	---------	------

[Write to the Help Desk](#)

[NCBI](#) | [NLM](#) | [NIH](#)

[Department of Health & Human Services](#)

[Freedom of Information Act](#) | [Disclaimer](#)

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:42 ; Search time 30.365 Seconds  
(without alignments)

1594.036 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106  
Sequence: 1 MELPVTHTLFLPLVFLTGLC.....GFPAHKKIPEEKREKLEQ 1167

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6106	100.0	1167	1	ITAG_HUMAN
2	2429	39.8	1189	1	ITAH_HUMAN
3	1894.5	31.0	1180	1	ITAI_RAT
4	1887	30.9	1151	1	ITAI_HUMAN
5	1756	28.8	1181	1	ITAI_HUMAN
6	1754	28.7	1170	1	ITAI_HUMAN
7	1744	28.6	1178	1	ITAI_HUMAN
8	1131.5	18.5	1162	1	ITAD_HUMAN
9	1105	18.1	1170	1	ITAD_HUMAN
10	1097	18.0	1163	1	ITAD_HUMAN
11	1079	17.7	1152	1	ITAM_MOUSE
12	1051	17.2	1163	1	ITAM_HUMAN
13	1022.5	16.7	1153	1	ITAM_MOUSE
14	953	15.6	1167	1	ITAE_MOUSE
15	932.5	15.3	1179	1	ITAE_HUMAN
16	738	12.1	1035	1	ITAE_HUMAN
17	690	11.3	1039	1	ITAE_MOUSE
18	689.5	11.3	1032	1	ITAE_MOUSE
19	684	11.2	1038	1	ITAE_MOUSE
20	674.5	11.0	1038	1	ITAE_MOUSE
21	629.5	10.3	1181	1	ITAE_HUMAN
22	602.5	9.9	1130	1	ITAE_HUMAN
23	600.5	9.8	1066	1	ITAE_HUMAN
24	596.5	9.8	1072	1	ITAE_HUMAN
25	596.5	9.8	1072	1	ITAE_HUMAN
26	586.5	9.6	1050	1	ITAE_HUMAN
27	586	9.6	1044	1	ITAE_HUMAN
28	584.5	9.6	1044	1	ITAE_HUMAN
29	581.5	9.5	1053	1	ITAE_HUMAN
30	576.5	9.4	1044	1	ITAE_HUMAN
31	574	9.4	1048	1	ITAE_HUMAN
32	573.5	9.4	1066	1	ITAE_HUMAN
33	572.5	9.4	1066	1	ITAE_HUMAN

34	566	9.3	1053	1	ITAE_MOUSE	P11688 mus musculus
35	563.5	9.2	1034	1	ITAV_CHICK	P26008 gallus galli
36	562	9.2	1049	1	ITAE_HUMAN	P08648 homo sapien
37	541	8.9	1146	1	ITAI_DROME	P24247 drosophila
38	540.5	8.9	1025	1	ITAB_HUMAN	P53708 homo sapien
39	534	8.7	1039	1	ITAB_HUMAN	P08514 homo sapien
40	504	8.3	1033	1	ITAB_MOUSE	P08514 homo sapien
41	485	7.9	1226	1	PAT2_MOUSE	P08514 homo sapien
42	475	7.8	1396	1	ITAE_DROME	P12080 drosophila
43	447.5	7.3	1000	1	ITAE_DROME	P12080 drosophila
44	445.5	7.3	1139	1	ITAE_DROME	P12080 drosophila
45	422.5	6.9	1115	1	ITAE_DROME	P12080 drosophila

## ALIGNMENTS

RESULT 1  
ITAG\_HUMAN STANDARD; PRT: 1167 AA.  
AC 075578; Q9UH28;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-10 precursor.  
GN ITGA10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Articular chondrocytes;  
RX MEDLINE=98352078; PubMed=9685391;  
RA Camper L., Hellman U., Lundgren-Akerlund E.;  
RT "Isolation, cloning, and sequence analysis of the integrin subunit  
RT alpha10, a beta1-associated collagen binding integrin expressed on  
RT chondrocytes.";  
RL J. Biol. Chem. 273:20383-20389 (1998).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart, and Endothelial cells;  
RX MEDLINE=20169197; PubMed=10702680;  
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,  
RA Wang S.-X., Langley R., Kirsans G.W.;  
RT "The integrin alpha10 subunit: expression pattern, partial gene  
RT structure, and chromosomal localization.";  
RL Cytogenet. Cell Genet. 87:238-244 (1999).  
CC - FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.  
CC - SUBUNIT: HETERO-DIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10  
CC ASSOCIATES WITH BETA-1.  
CC - SUBCELLULAR LOCATION: Type I membrane protein.  
CC - TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN  
CC MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.  
CC - DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC - SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC - SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC - SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF074015; AAC1952.1; -  
CC EMBL: AF112345; AAP21944.1; -  
CC EMBL: AF172723; AAF61638.1; -  
CC HSSP: P17301; IAOX.  
CC Genew: HGNC:6135; ITGA10.

[illegible]

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal heart, and Osteoblast;  
 RX MEDLINE=99417678; PubMed=10466209;  
 RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,  
 RA Wang S.-X., Morris C.M., Kirsansen G.W.;  
 RT "Cloning, sequence analysis, and chromosomal localization of the novel  
 RT human integrin alpha1 subunit (ITGA11).";  
 RL Genomics 60:179-187(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal muscle, and Uterus;  
 RX MEDLINE=99395147; PubMed=10464311;  
 RA Velling T., Kuesche-Gullberg M., Sejersten T., Gullberg D.;  
 RT "cDNA Cloning and Chromosomal Localization of Human Alpha(11)  
 RT Integrin. A collagen-binding, 1 domain-containing, beta(1)-associated  
 RT integrin alpha-chain present in muscle tissues.";  
 RL J. Biol. Chem. 274:25735-25742(1999).  
 RN [3]  
 RP SEQUENCE OF 954-1188 FROM N.A.  
 RC TISSUE=Fibroblast;  
 RA Andreu N., Escivill X., Escarceller M., Sumoy L.;  
 RL Submitted (Jun-2000) to the EMBL/Genbank/DBD databases.  
 CC -1- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.  
 CC -1- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND  
 CC HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO  
 CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO  
 CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS,  
 CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN  
 CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.  
 CC -1- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING  
 CC FETAL MUSCLE CELLS (IN VITRO).  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF109681; AAF01258.1; -;  
 DR EMBL, AF137378; AAD51919.2; -;  
 DR EMBL, AL359064; CAB94392.1; -;  
 DR HSSP; P17301; IAOX.  
 DR GenSeq; HGNC:6136; ITGA11.  
 DR MIM: 604789; -;  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VFMA\_A.  
 DR Pfam; PF00092; vfa; 1.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VFMADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; vfa; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.  
 DR PROSITE; PS50234; VFMA; 1.  
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal; Repeat; Calcium; Magnesium; Polymorphism.  
 FT SIGNAL 1 22  
 FT CHAIN 23 1189 INTEGRIN ALPHA-11.  
 FT DOMAIN 23 1142 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1143 1165 POTENTIAL.

FT DOMAIN 1166 1189 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 38 94 FG-GAP 1.  
 FT REPEAT 102 163 FG-GAP 2.  
 FT DOMAIN 167 345 VFMA.  
 FT REPEAT 359 420 FG-GAP 3.  
 FT REPEAT 422 475 FG-GAP 4.  
 FT REPEAT 477 537 FG-GAP 5.  
 FT REPEAT 539 598 FG-GAP 6.  
 FT REPEAT 601 653 FG-GAP 7.  
 FT DOMAIN 1154 1162 POLY-LEU.  
 FT DOMAIN 1174 1177 POLY-ARG.  
 FT CA\_BIND 468 496 POTENTIAL.  
 FT CA\_BIND 551 559 POTENTIAL.  
 FT CA\_BIND 613 621 POTENTIAL.  
 FT DISULFID 76 83 BY SIMILARITY.  
 FT DISULFID 121 139 POTENTIAL.  
 FT DISULFID 129 159 POTENTIAL.  
 FT DISULFID 659 668 BY SIMILARITY.  
 FT DISULFID 674 729 BY SIMILARITY.  
 FT DISULFID 781 787 BY SIMILARITY.  
 FT DISULFID 881 893 BY SIMILARITY.  
 FT CARBOHYD 82 82 N-LINKED (GLCNAC...)  
 FT CARBOHYD 95 95 N-LINKED (GLCNAC...)  
 FT CARBOHYD 291 291 N-LINKED (GLCNAC...)  
 FT CARBOHYD 311 331 N-LINKED (GLCNAC...)  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...)  
 FT CARBOHYD 449 449 N-LINKED (GLCNAC...)  
 FT CARBOHYD 462 462 N-LINKED (GLCNAC...)  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC...)  
 FT CARBOHYD 642 642 N-LINKED (GLCNAC...)  
 FT CARBOHYD 694 694 N-LINKED (GLCNAC...)  
 FT CARBOHYD 857 857 N-LINKED (GLCNAC...)  
 FT CARBOHYD 894 894 N-LINKED (GLCNAC...)  
 FT CARBOHYD 973 973 N-LINKED (GLCNAC...)  
 FT CARBOHYD 1032 1032 N-LINKED (GLCNAC...)  
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC...)  
 FT VARIANT 433  
 FT VARIANT 524  
 FT VARIANT 524  
 FT VARIANT 972  
 FT VARIANT 1003  
 FT VARIANT 1030  
 FT VARIANT 1094  
 SQ SEQUENCE 1189 AA; 133609 MW; 60303C08A4A4CD52 CRC64;  
 Query Match 39.8%; Score 2429; DB 1; Length 1189;  
 Best Local Similarity 43.1%; Pred. No. 5.7e-167; Indels 46; Gaps 16;  
 Matches 516; Conservative 213; Mismatches 422;  
 QY 1 MELPFVTHFLPLVFLTGLCSFNLDEHPRLPFGPPAEERGVSYLQHVGGGQRMVLVGA 60  
 DB 1 MDLPRLGVVAVMALSLMPGFDTFPMNDRKPRVIESRRAFFGYTQGDIGSKMLVYGA 60  
 QY 61 PWDGSGDRRGDYRCPLVGAHANAPCAKGLHGDYOLGNSSHPAVNMLGMSLLETDGCG 120  
 DB 61 PLENGYOKTGDVYKCPV---IHGCTKLNIGRVLTSNVSERKMKRMLGLSLATNPKN 117  
 QY 121 FMACAPLMSRAGSSVFSSGICARVDASFQOGSLAPTAOBCPTVMADVIVYLDGNSITYP 180  
 DB 118 FLACSPLMSSHEGSSYTTGMCISVNSFRSKVAPLQRCQTYMDIVIVYLDGNSITYP 177  
 QY 181 MSEVOTFLRLRVGLKFLIDPEQIOVGLVGYGSSPVHEWSLADGRTKEEYVRAKMLSRREG 240  
 DB 178 WVEVGHFLINILTKRYIGPQIOGVGVGDVHVEFLINDYRSKVDVVEASHIEGRG 237  
 QY 241 RETKRAQIMVACTEGESQSGRPEARLLVVTVDGSHGEBELPAALKACEAGRVTRY 300  
 DB 238 TETRTAFIEIFARSEAFQK--GGRKAKKVMIVITDGHSDSPDLEKVIQOOSERNVTRY 295

QY 301 GIALGVILRRQDPSSFLREIRTIASDPDERFFVNVTEAALTDIVDALGRIFGLEGS 360  
 DB 296 AVAVLGVNRRGINPETFLEIKYIASDPDCKFFNVTEAALAKDIVDALGDSIFLEEGT 355  
 QY 361 HAENESSFGLMSQIGPSTRLKDGILFGVNAVYDMGCVLMLEGRHLPFPMALDEDF 420  
 DB 356 N-KNETSPFGLMSQIFSSSHVEDGVLLGAVGADYMNAGVALKETSAKGVPLFEESYLKEF 414  
 QY 421 PPAIOWHAAYLVGVSSMLRGSRRLFLSGAPRRHKGKVIATQLKKDGVRAVAGSLOOE 480  
 DB 415 PEELKNGHALGVTVTSVSSVSGRGKRVYVAGAPRNHTGKYLFTMNNNSLTIHQMKRG 474  
 QY 481 QIGSYGSELCPDTRDGTDDVLLVAAPMLCPQNKETGRVYVYVVGQOSLTLTQIG 540  
 DB 475 QIGSYGSEITSDIDGDDVTLVGAPEYFN-EGREKGVYVEL-RQNRVYVNGTLX 532  
 QY 541 PEPP-ODARFGPMALPDUNOGFADVAVGAPLEDOHQALYLTGTSQGVAPHPAQR 599  
 DB 533 DSHSYGNARFGSIAVVDLNDQSYNDVVVGALEENHAGAIYIFGFRSILKTEKOR 592  
 QY 600 AASMHALSYFGRVDGRLLDGDLLVAVAGAGAILSSRPVHLTPSLEVTPO 659  
 DB 593 TASELTGLQYFGCSIHGQDLNEDGLIDAVAGLGNVAVLWSRPVQVNASLHPEPSKI 652  
 QY 660 SVVQDRCRRGCAVCLTAALCFQVTSRTPGRMDHOFYMEFTASLDEMTAGARAAPDGS 719  
 DB 653 NIFNRCKSGRATLAAFLCTPFLAPHPOTTVGIRYVNTMERRTPPAHLDEG 712  
 QY 720 QRLSPRLRLSVGNVTCQOLHFLVLDTSYLPRVALVTFPALDNTKPGVNLGSPSI 779  
 DB 713 DRFTNPAVLLSSQCELCERINFVLDADYKVPFESVSESLDDPH-GPMLDDGPTTL 771  
 QY 780 OKVPSKXCGPNECVTDVLQVNDI-----RGSXK-----APFV 817  
 DB 772 RVSPFWNGCNEDHCVPDLVLDARSDDLPTAMEYCORVLRKPAQDCSANTLSFTTVFI 831  
 QY 818 RGRGRKVLSTLENKENAVNTSLIIFSRNLHLASLTPORESPIKVECAAPS-AHAR 875  
 DB 832 ESTROGVAVAEATLENNGENAVSVTLNISOSANQPASLICKEDSDSIECVNERLQ 891  
 QY 876 LCSVGHVQTCAKVFLLEFEFSSCSLSQVFGKLTASSDSLEKNGTLOENTAOISAV 935  
 DB 892 VCNVSYPPFRACKVAFRLDFEFSKSIFLHLEIEIAGSDSNERSTKEDNVAAPRFLH 951  
 QY 936 OVEPHLLFSESTLHRYEVHVGTLR-VGPGPEFTTLRVONLGVVSGLIISALPA 993  
 DB 952 KTEADVLFTRSSLSHYEVKLSLSERYDGIQPFSCIFRIQLGLFPIHGLWMTTPI 1011  
 QY 994 VAHGVNVLSSQVITN--NASC-IVONLTERPPGPVHPEELOHTNRNLGNSNTQCOVR 1049  
 DB 1012 ATPSGRRLKLRLFLDEVANTSCNIMGNSTEVPRPVE-EDLRARPOLNHSDDVVSIN 1070  
 QY 1050 CHIGOLAKTEVSVGLRLVHNEFPFRACKSLTVVSTELGTBESVUQLTEASWSS 1109  
 DB 1071 CNI-RLVPOEINFLHGLNMLRSLKALKKXKIMVNAALQOFHSPTFFREEDPSROI 1129  
 QY 1110 LLEVOTRPLILSLMLIGSVLGLLLALVLCMKLGFPAKRIPEEKREKXE 1166  
 DB 1130 VBEISKQEDMVPIWITVSTLGLLLALLVLAALKGFPSAR---RRREPGD 1182

RESULT 3  
 ID ITAT RAT STANDARD; PRT; 1180 AA.  
 AC P18614;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-1 precursor (laminin and collagen receptor) (VLA-1)  
 DE (CD49a).  
 GN ITGAL.  
 OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90338125; PubMed=2380249;  
 RA Ignatius M.J., Houde T.H., Tawil J.W., Barton A.,  
 RA Esch F., Carbonetto S., Reichardt L.F.;  
 RT "Molecular cloning of the rat integrin alpha 1-subunit: a receptor  
 RT for laminin and collagen-";  
 RL J. Cell Biol. 111:709-720(1990).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.  
 RX MEDLINE=99311397; PubMed=10386626;  
 RA Nolte M., Pepinsky R.B., Vanyaminov S.Y., Kotellanev V.,  
 RA Gotwals P.J., Karpusas M.;  
 RT "Crystal structure of the alphabeta1 integrin I-domain: insights into  
 RT integrin I-domain function";  
 RL FEBS Lett. 452:379-385(1999).  
 CC -1- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND  
 CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-  
 CC E-R IN COLLAGEN.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTRAINS 1 VWFA DOMAIN.  
 CC -1- SIMILARITY: CONTRAINS 7 FG-GAP REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/annouce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X52140; CAA36384.1; -.  
 DR PIR: A35854; A35854.  
 DR HSP: P17301; IAOX.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR InterPro: IPR002035; VWF\_A.  
 DR Pfam: PF00092; vwa; 1.  
 DR Pfam: PF00357; Integrin\_A; 1.  
 DR Pfam: PF01839; FG-GAP\_5.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 DR SMART: SM00191; Int\_alpha; 5.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE: PS50234; VWFA; 1.  
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal; Repeat; Calcium; Magnesium.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1180 INTEGRIN ALPHA-1.  
 FT DOMAIN 29 1142 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1143 1165 POTENTIAL.  
 FT DOMAIN 1166 1180 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 44 103 FG-GAP 1.  
 FT REPEAT ? ? FG-GAP 2.  
 FT DOMAIN 175 388 VWFA.  
 FT REPEAT 377 432 FG-GAP 3.  
 FT REPEAT 433 484 FG-GAP 4.  
 FT REPEAT 485 565 FG-GAP 5.  
 FT REPEAT 567 629 FG-GAP 6.  
 FT REPEAT 629 681 FG-GAP 7.  
 FT CA\_BIND 497 505 POTENTIAL.  
 FT CA\_BIND 579 587 POTENTIAL.  
 FT CA\_BIND 641 649 POTENTIAL.  
 FT SITE 1168 1172 GPRK MOTIF.  
 FT DISULFID 82 92 BY SIMILARITY.  
 FT DISULFID 687 696 BY SIMILARITY.

```

FT DISULFID 702 755 BY SIMILARITY.
FT DISULFID 807 813 BY SIMILARITY.
FT DISULFID 877 885 BY SIMILARITY.
FT DISULFID 1029 1062 BY SIMILARITY.
FT DISULFID 1066 1073 BY SIMILARITY.
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 459 459 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1007 1007 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1103 1103 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1180 AA; 130808 MW; 855DA2BE02362EE4 CRC64;

```

```

Query Match 31.0%; Score 1894.5; DB 1; Length 1180;
Best Local Similarity 36.5%; Pred. No. 2, 1e-128;
Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;

```

```

QY 13 LVFLTGLCSPPNLDENHPRLPFGPPAEFGYVLQHVGGGQRMVLVGAERDPSGDRGD 72
DB 19 LTVIILGFCVFNVDVNSMFSFGPVEDMFGYTVQCEVNEGKMWLIGSLVPGQPARATGD 78
QY 73 VYRCVGGGHNAPCAKKGHLDYOLGSSHPAV-----NHLGMSLFTDGGGPMACPL 127
DB 79 VYKCPVGRBRAMPVCKLDLP---VNTSLPNVTEIKENMTFG-STLVNPNNGFLACGLP 133
QY 128 WSRACGSSVSSGICARVDASFQPOGSLAPTACPTVMVDVIVDGSNSIYPMSEVQTF 187
DB 134 VYRCGHLHTYTIGICSDVSPFTQVNSFAP-VQECSTQIDYIVDGSISITPMSSVIAF 192
QY 188 LRLVGLKFLIDPQIQVGLVQYGESFVHWSLGDFTKEEVVRAAKNLSRRRGRETQAQ 247
DB 193 LINDLKRMIDIGKQOVGIVQYGENVTHEFNINKYSSTEBEVLVAAANKIGRQGLQTMPL 252
QY 248 AIMVACTEFGSOSHGRPEAAALLVVTTCGSHDDEELPAALKAEEAGVNTYGVAVLGH 307
DB 253 GIDTARKEFTARGARRGVKKVMVLTVDGESHDNVRLKQVQIDCEBENIQFSSIALIGH 312
QY 308 YLRRODSSFLREIRTIASDDERFFVNTDEAALTIDVDAIGRIFGLGSHANESS 367
DB 313 YNRGNLSTKEFVEELIKSISTEPTEKHFVNSDELALVTIVKALGRIPALEXTAQAAS 372
QY 368 FGLEMSQIGFSTHRLKDGILFGMVAGAYVWGSSVLMLEGHRLFPRMALEDEFPALONH 427
DB 373 FEMEMSQIGFSAHYQSDWMLGAVGAYDNGVTVMQKANOVIYPHNTTFQTE--PAKME 430
QY 428 --AALVGSVSSMLRGGRRLFLSGAPRRHKGXIAFOLKDGAVRAVQSIQSGIQSY 485
DB 431 PLASVIGYVNSATIPGD-VLYIAGOPRNHNGOVVITM--BDGINILQTLTGSGIQSY 488
QY 486 FGSSEICPLDTRDGTDLVLAAMPFLGPNKQETGRVYVLYVGOOSLTLQSTLGP--- 541
DB 489 FGSVLTIDIDKDYTLDLVGAAPMVGTEKEQGVYVYAN-NQTRFQMSLEPIKOT 547
QY 542 -----EPQDARFGFAGMALPDINODGFADVAVGAFLEDGHOGALYLYHG 586

```

```

DB 548 CSSLKDNSCTKENKNEBCGARFGTALAAYKDLNVDFGNVYVIGAPLEDHAGAVYIYHG 607
QY 587 TQSGVRPAPRAORIAAASMPHALSYRGVSDRLDDGDLDVAVAGAGAAIILSRPIV 646
DB 608 SKGTRERAVQAIRPBGSGCKTLKFFGOSIHGBMDLNGCLTDVYIIGGAGALFVARVA 667
QY 647 HLTPELVTPQAIISVVRDCCRROGAAYCLTALCFQYTSRTPGRMDHOFYRFTASDE 706
DB 668 VVKVTMNEPMPKVNIOKKNCGVEGETEYCINATMCFHKLKSKEDSIYEAOLQYRVTLDS 727
QY 707 WTAGRAAFPDGSGORLSPRLPLSYGVANTCEQLHNVHVDTDYLRPVALYTFALDNTTK 766
DB 728 LRQISRSEFSGTOEKIOR--NITYRESECRHSGFYMDKHDFODSVAVTIDF--NLTJ 782
QY 767 P--GPVINEGSPTSIOKLVPSKDCGPNCEVTDLVQVNDINGSRAPFVGRGRKV 824
DB 783 PENGEVLDALPNSVHEIIPRAKDCGKERCIISDITLVST---TEKSLIVSQHDF 838
QY 825 LVSTILENRKENAYNTSLIIFSRNHLASLTP-QRESPIVECAPSAHARLCSVGHV 883
DB 839 NVSLTVKXKGDVAVTRTVVQHSPLVPSGIEIQKDS-----CESNQNTICRGVYF 891
QY 884 FQTKAKTFLLEPFESGSLSYVFGKLTASSDSIERNGTIOENTAOISAYIQYEPHLLF 943
DB 892 LRAGETVTFKIIIFQNTSHLSENAIIMLSATSDSEBPLESLNDRVENVISIPKVEVGLOF 951
QY 944 SSESTLHRYEVPYGTLP-----VGPPEKRTKTLRYONLCGYVSGLIISALLP-AVAH 996
DB 952 YSSASEHHISVAANETIPEFINSTEDIGENINVFYTKRGHPFAPLQSLISFPNLTRD 1011
QY 997 GGNVFLSLSOVITNNASCIYVNLTEPPG-----PPVHPEELQHTNRINGSNTOCOVVC 1050
DB 1012 GYPVLVYIGWSSSDVNRPRSLDEPFGINGKMTKISKEVLKRGTIQDCSSITGVATI 1071
QY 1051 HLGQLAKC-TRVSGCLRLVNEPFRAPKFSLYVSTFELCTEBSVYLOLTSRWSBS 1109
DB 1072 TCSLPSDLSQVNVSL-LWKPTFIKRAHFSLSNLTLEGEKSENS-LTSSSRKREL 1128
QY 1110 LLEVQOT-RPLISLWILIGSVLGLLALLVFCMLWLGFAHKKIPEEKREE 1163
DB 1129 AIQISKDGLPGRVPLWVILLSAFGLLMLLILALMKIGFF---KRLPKKMEK 1180

RESULT 4
ITAL HUMAN STANDARD. PRT; 1151 AA.
AC P56199;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
GN ITGA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 11;
RP SEQUENCE FROM N.A.
RX MEDLINE=93155124; Pubmed=8428973;
RA Briesewitz R.; Epstein M.R.; Marcantonio E.E.;
RT "Expression of native and truncated forms of the human integrin alpha
  1 subunit."
RL J. Biol. Chem. 268:2989-2996(1993).
CC -!- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC E-R IN COLLAGEN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFPA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VFPA DOMAIN.

```

CC - - SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC - - DATABASE: NAME=PROW; NOTE=CD guide CD49a entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd49a.htm".  
 DR HSSP; P17301; IAOX.  
 DR Genew; HGNC:6134; ITGAL.  
 DR MIM; 192968; .  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR InterPro: IPR002035; VWF\_A.  
 DR Pfam; PF00092; vwa; 1.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00191; Int alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS0234; WFA; 1.  
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Repeat; Calcium; Magnesium.  
 FT DOMAIN 1 1113 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1114 1136 POTENTIAL.  
 FT DOMAIN 1137 1151 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 16 75 FG-GAP 1.  
 FT REPEAT ? 75 FG-GAP 2.  
 FT REPEAT 147 360 WFA.  
 FT DOMAIN 349 404 FG-GAP 3.  
 FT REPEAT 405 457 FG-GAP 4.  
 FT REPEAT 459 520 FG-GAP 5.  
 FT REPEAT 540 599 FG-GAP 6.  
 FT REPEAT 602 654 FG-GAP 7.  
 FT CA\_BIND 470 478 POTENTIAL.  
 FT CA\_BIND 552 560 POTENTIAL.  
 FT CA\_BIND 614 622 POTENTIAL.  
 FT SITE 1139 1142 GEFKR MOTIF.  
 FT DISULFID 54 64 BY SIMILARITY.  
 FT DISULFID 660 669 BY SIMILARITY.  
 FT DISULFID 675 728 BY SIMILARITY.  
 FT DISULFID 780 786 BY SIMILARITY.  
 FT DISULFID 850 858 BY SIMILARITY.  
 FT DISULFID 1002 1034 BY SIMILARITY.  
 FT DISULFID 1037 1044 BY SIMILARITY.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 812 812 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 855 855 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1085 1085 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1151 AA; 127837 MW; 683F3C1AABF52808 CRC64;

Query Match 30.9%; Score 1887; DB 1; Length 1151;  
 Best Local Similarity 36.3%; Pred. No. 7e-128;  
 Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;  
 23, FNIDENHRLFPGRPEAFGYSVLQHVGGGQRMVLVGAWDPGSGDRGDDVYRCVPGAH 82

Db 1 FNVDFKSNMTSGPVEDMFGTVQYQYENEBEGKMWLIGSLVVGQPKNRIGDYKCPVGRGE 60  
 QY 83 NAFCKAGHLDYQOLGNSSHPAV----NMHLGMSLLETDGGGFMACAPLMSRACGSSVF 137  
 Db 61 SLPCVKLDP-----VNTSIPNVTVEKEMMTG-STLVTPNPGGFLACCPLYAYRGCHLHY 115  
 QY 138 SSGICARVDASFPQCGSLAPTAQRCPYMDVYIVLDGNSNIYPMSGYOTFRLRYGKLF 197  
 Db 116 TTGICSDVSPFTFOVNSIAP-VQECSTQDVIYLDGNSNIYPMDSVTAFLNDLKRD 174  
 QY 198 DPEQIQVLYOYGESPVHEWSLGDPRIXEEVYRAAKNLSRREGRETQAQIMVACTGCF 257  
 Db 175 GKQIQVQIVQYGENVTHEFNLNKYSSTEEVLYAAKKIVQGGQRTMTALGDTARKEAF 234  
 QY 258 SOSHGCRDEARLVLVYTDGSHDEELPALKACEARVTRYCAVLGHILRRDRSS 317  
 Db 235 TEARGARGVKKWVIYTDGSHDHRLLKVIQDEDEENIQRFSLAIGSYNRGLSTEK 294  
 QY 318 FLREIRIASDPDERFFPNVTDEALTDIYDALGRIFGLSGSHANESSFGLMSQIGF 377  
 Db 295 FVEEIKSIASEPTGHPFNVSDELALVTITLGRITALEATADQASAFSEMESQIGF 354  
 QY 378 STHRLKDGILFGMVGAYDMGSLWLBEGHRLFPRLALDEFPALONHAAYLGYSVS 437  
 Db 355 SAHYSQDWVMVGAVAYDMNGVTVMQKASQIIIPRNTFNVESTKKNPLASYLGTVNS 414  
 QY 438 MLIRGRLLFLSAGAFRRHGRKVIAFOIKQGAARVAVSLOGEOIQSGFSGELCPDIDR 497  
 Db 415 ATASSGDVLYIAGQRYVHTQVLIYRM-EDGNKIILDTLGEQIGSYFGSILTTTIDK 473  
 QY 498 DDTTLVLVAAPMPFGPONKETGRVYVYLVGQSLTLTQGLQP----- 541  
 Db 474 DSNTIILVGAAPMWTGTEKEQKYVYAL-NQRFYQMSLEPRKOTCCSSROHNSCTT 532  
 QY 542 ---BPPODARFGFAMGALPDINODGADAVAGAPLEDGQALYLYHGTQSGVRPAPQR 598  
 Db 533 ENKNEPCARGRTAIAVKDINLDGFNDIVIGAPLEDHGAIVYHSGSKTIRKEYAQR 592  
 QY 599 IAAASMPALSYFGSVSGRLDLOGDDLVDAVAGQAAILLSRPYHLPSLEVTQA 658  
 Db 593 IPSGGDGKTLRFPGGSIHGEMDNDGLTDVITGGLGALFLWSDDAVAVVYTNMFEENK 652  
 QY 659 ISVORDCRRRGOEAVCTLAACFOVTSRTGPMHOFYMFETASLDEWTAGARAAPFGS 718  
 Db 653 VNIOQKCHMEGKEVVCINATVCEPVKLSKEDTITYEDLQRYVLDSLRQISRFSGT 712  
 QY 719 GQRLSRRLRLSVGNVTCQQLHFNVLDTSDYLRLPVALTVPALDNTYKP--GPVINEGSP 776  
 Db 713 QERKVR--NITVRKSECTKHSFYMLDGHDFQDSVRITLDF--NLTDPENGPVLDDSLP 767  
 QY 777 TSIOQLVPSDCGCGDNECVTDVLYQVMMDIRGSKAKAFVVRGGRKYLVTLEENREN 836  
 Db 768 NSVHEYIIPADCGKKEKICISDLHV---ATTEKDLIYRSQNDKNVSLTVNNTDS 823  
 QY 837 AYNTLSLIIIFSRNLILASTPQRESPIKYVECAASAHARLSVGHVPVQATKATFLEF 896  
 Db 824 AYNTTIYHSPNLVFSGI-----EAIQKDS--ESNHNITCKYGPFLRBMGMVTFKILF 877  
 QY 897 EFSSGSLISQYFGKLTASDSLERNLTQENTQTSAYIOYEPHLLPSSSESTLHYEYHP 956  
 Db 878 QENTSYLMENVTIYLSATSDSEEPETLSDVNVNISIPVKEYEVQIFYSASEYHISIAA 937  
 QY 957 YGTLP-----VGPEPEKTLRQVNLGCVYVSGILITALLPVAHGN---YFLSLQV 1007  
 Db 938 NETVEVINSTEDIGNEINIFYLIRKSSFPMBELKISIFPNMTSNQYPVLYPTGLSS- 996  
 QY 1008 ITNNASCIVQNLTP-----PGPVHPELOHTRLNGSNQOCVAVCHLQGLAKTEV 1061  
 Db 997 -SENANCPHFEDFINSKMTTSDHLKRGITLDONCKPATITCNLTS-SDIQV 1054  
 QY 1062 SVGLRLVHNEFFRRAKFKSLTVSTFELGTEEGSVLQLTASRWSLSLLEVQOT-RPIL 1120



DB 1055 NWSL--ILMKPFIKYSFSSNLTITNGELRSENAS-LVUSSNOKRELAIOISKDGLPCR 1111  
QY 1121 ISLWILIGSVLGLLALLVFLCMLGCPFAHKKIPEEKREB 1163  
DB 1112 VPLWVILTSAPFGLLMLLMLLWKIGF---KRLKKKMK 1151

RESULT 5  
ID T1A2 HUMAN STANDARD; PRT: 1181 AA.  
AC P17301;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)  
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).  
GN ITGA2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. AND SEQUENCE OF 30-44.  
RC TISSUE=Endothelial cells;  
RX MEDLINE=89308879; PubMed=2545729;  
RA Takada Y., Hemler M.E.;  
RT "The primary structure of the VLA-2/collagen receptor alpha 2 subunit  
RT (platelet GPIa): homology to other integrins and the presence of a  
RT possible collagen-binding domain.";  
RL J. Cell Biol. 109:397-407(1989).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.  
RX MEDLINE=98019223; PubMed=9353312;  
RA Emsley J., King S.L., Bergelson J.M., Liddington R.C.;  
RT "Crystal structure of the I domain from integrin alphabeta1.";  
RL J. Biol. Chem. 272:28512-28517(1997).  
RN [3]  
RP VARIANT HPA-5 (BR).  
RX MEDLINE=94043762; PubMed=7901236;  
RA Santoso S., Kalb R., Walke M., Kiefel V., Mueller-Eckhardt C.,  
RT Newman P.J.;  
RT "The human platelet alpha2B integrin Br(a) and Brb are associated with a  
RT single amino acid polymorphism on glycoprotein Ia (integrin subunit  
RT alpha 2).";  
RL J. Clin. Invest. 92:2427-2432(1993).  
RN [4]  
RP VARIANT GLU-534.  
RX MEDLINE=20206009; PubMed=10744142;  
RA Kroell H., Gardemann A., Fechter A., Haberboesch W., Santoso S.;  
RT "The impact of the glycoprotein Ia collagen receptor subunit A1648G  
RT gene polymorphism on coronary artery disease and acute myocardial  
RT infarction.";  
RL Thromb. Haemost. 83:392-396(2000).  
CC -I- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,  
CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT  
CC RECOGNIZES THE POLYLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN  
CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER  
CC CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE  
CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
CC EXTRACELLULAR MATRIX.  
CC -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
CC ASSOCIATES WITH BETA-1.  
CC -I- SUBCELLULAR LOCATION: TYPE I membrane protein.  
CC -I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -I- POLYMORPHISM: POSITION 534 IS ASSOCIATED WITH PLATELET-SPECIFIC  
CC ALLOANTIGEN HPA-5 (BR). HPA-5A/BR(A) HAS LYS-534 AND HPA-5B/BR(B)  
CC HAS GLU-534. HPA-5B IS INVOLVED IN NEONATAL ALLOIMMUNE  
CC THROMBOCYTOPENIA (NAT OR NATP). THE K534E POLYMORPHISM MAY PLAY A  
CC ROLE IN CORONARY ARTERY DISEASE (CAD).  
CC -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -I- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.

CC -I- DATABASE: NAME=PROW; NOTE=CD guide CD49b entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49b.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL, X17033; CA34894.1; -  
CC PIR, A33998; A33998.  
CC PDB, IAOX; 25-NOV-98.  
CC Genew; HGNC:6137; ITGA2.  
CC MIM; 192974; -  
CC InterPro; IPR000413; Integrin\_alpha.  
CC InterPro; IPR002035; VWFA.  
CC PIR, PF00092; VWFA; 1.  
CC Pfam; PF00357; Integrin\_A; 1.  
CC Pfam; PF01839; FG-GAP; 5.  
CC PRINTS; PR00453; VWFADOMAIN.  
CC SMART; SM00391; Int\_alpha; 5.  
CC SMART; SM00327; VWFA; 1.  
CC PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
CC PROSITE; PS00234; VWFA; 1.  
CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
CC Platelet; Signal; Repeat; Calcium; Magnesium; Polymorphism;  
CC 3D-structure.  
CC SIGNAL 1 29  
CC CHAIN 30 1181  
CC DOMAIN 30 1132  
CC TRANSMEM 1133 1154  
CC DOMAIN 1155 1181  
CC REPEAT 45 103  
CC REPEAT ? 378  
CC DOMAIN 188 433  
CC REPEAT 378 433  
CC REPEAT 434 486  
CC REPEAT 488 549  
CC REPEAT 551 610  
CC REPEAT 615 667  
CC CA\_BIND 499 507  
CC CA\_BIND 563 571  
CC CA\_BIND 627 635  
CC SITE 1157 1161  
CC DISULFID 83 92  
CC DISULFID 680 737  
CC DISULFID 789 795  
CC DISULFID 865 876  
CC DISULFID 1019 1050  
CC DISULFID 1055 1060  
CC CARBOHYD 105 105  
CC CARBOHYD 112 112  
CC CARBOHYD 343 343  
CC CARBOHYD 432 433  
CC CARBOHYD 460 460  
CC CARBOHYD 475 475  
CC CARBOHYD 699 699  
CC CARBOHYD 1057 1057  
CC CARBOHYD 1074 1074  
CC CARBOHYD 1081 1081  
CC VARIANT 534 534  
CC SQ SEQUENCE 1181 AA; 129295 MW; 7E1B7ED968A94070 CRC64;  
Query Match 28.8%; Score 1756; DB 1; Length 1181;  
Best Local Similarity 34.3%; Pred. No. 2.1e-118;  
Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;  
QY 11 LPLVFLTL-----CSFPLNDEHHRLLPFGPEAEAGYGVLDVHGVGCGQWMLVGAPWD 63  
DB 11 LPLLVLTALSGGIINCCIAVNVGLPEAKIFSGPSSEQFGYAVVQGFINDKGNMLLVGSPWS 70

QY 64 GPSGDRGDVYRCPVGGAHNPACAKGHT--GDYQLGNSHPAVNMHLSLLETDGSGGM 122  
 DB 71 GFENMGDYKCPV-DLSTATCEKLNLTOSTSI PNTBEMKTMWSLGLITRMGIGGFL 129  
 QY 123 ACAPLMSRACGSSVFSGTCARVDASFQOGSIAPTAQRCPYTMVYVILDGNSIYPPS 182  
 DB 130 TCGPLMAOOCNGNYTTTGVCSDISPFOLSASFSPATQPCPSLIDVWVCDESNSIYPPMD 189  
 QY 183 EYQTFRLVGVKFLIDPEQIYGVLYOGVSPVHMSLGDRTKEEVYRAKULSREGRE 242  
 DB 190 AVKNPLEKPYOGIDIGPTKIQVGLIQANNPRVFNPLNTYKTEEMIVATSGTSQYGGL 249  
 QY 243 TKTAQAIMVACTEFGFOSHGPREAARLLVVTVDGSHDEELPAALKACEAGRYRYGI 302  
 DB 250 TMTFGAIQYARKAYAAASGGRSAIKWVWVVDGSHOSMKAKAIDOCNHNILRFGI 309  
 QY 303 AVUGHYLRORPSSFLREIRTIASDPDERFFNVVDEALTDIYDALGDRIFLEGSHA 362  
 DB 310 AVUGYLNRLNADLTKNLIKAIASIPTERYFFNVVDEALTEKAGTLGEOIFSIGET-V 368  
 QY 363 ENSSFGLEMSQIGFST--HRLKDGLFGWVGVDMGGSVLM--LEGHLLFPRMALDEB 419  
 DB 369 QGGDNFQWMSQVFSADYSSQNDIIMLGAVGAFGMSGITVQKTSHGHLFP---KQA 423  
 QY 420 FFPALO--NHAAYLGYSVSSMLRGGRRLFLSGAPFRHKGXVIAFOLKKDGAVVAQGL 477  
 DB 424 PQDILDRHSSSYLGYSVAA--ISTGSETHVAPAPANTGQVLYSVNENGITVIOAH 482  
 QY 478 QGEQIISYSGSELCPDLDRTDGTVDVLAAPMFLPQNKETRYVY---LVGQOGL 532  
 DB 483 RGQOISYSGSVYCSVDVDKDITDVLVGAPEWMSDLKKEBERVYLFTIKKQILCOHOF 542  
 QY 533 LTLQGTLOPEPPDARFGFAMGLPDLNDGFAADVAVGAPLEGGHOGALYLVHGOSGR 592  
 DB 543 --LEG--PEGIENTFGSHIALSDINDGFNDVIVGSELEQNSGAYITNGGCTIR 597  
 QY 593 PHPAORIAA--SMPLASYFGSRVDGRDLDDDDVDVAVAGOGAAILLSRPVHLTP 650  
 DB 598 TKYSQKILGSDGAFRSHLQYFGRSLDGYGLNDSTIDVSIAGFGVQVLMGSIADVAI 657  
 QY 651 SLEVTFOAISVORDCRRRQGEAVCTALCFQVTRTGRMDHQYVRFYSLDEWTAG 710  
 DB 658 EASFPEKILTVKNKNAQ-----IILKICFSAKER-PTKONNAVAIYVNTILDA---- 704  
 QY 711 ABAAPFGSGORLSRRL-----RLSGNV-----TCQGLRHVLDTSQVLRVALLTV 758  
 DB 705 ----DGFSSRYTSRGLFKENNERCLOKNNVNVQAOSCEPHIITYQEPSDVVNSLDLRVD 759  
 QY 759 FALDNTTKPG--PVLNEGSPITSIOKLVPFSKDCGPNCEVTDLVLOVNMDIRGSRKAPV 816  
 DB 760 ISLEN---FGTSPALEAVSETAKVFSIPFHKOCGEGLCISDVLVDVR-QIAPAQOPRI 815  
 QY 817 VNGGRKVLVSTLLENKRENAVNTSIIIFSRNLHLASLTPORESI---KYEC-AAPSA 872  
 DB 816 VANONRRLTFSSVTLKNNKRESAVNTGIVVFSENLFFASFS---LPVDSTEVTCQVAAQO 871  
 QY 873 HARLGVGHVPOTGAKTFLLEFFRSCSLSQVFGKLTASDSLERGTIOENNAQTS 932  
 DB 872 KSVACVGVPALRREGQVFTTFINFDNLONLQOASLSFOALSESSEENKA--DNLVNIK 929  
 QY 933 AYIOYEPHLLFSSSESTLHRYEVPYTLP-----VGPPEFKTLTRVONLCYVVSGLI 986  
 DB 930 IFLVDABEHLTRSTININFEIISDQNVPSIYHSPEDVCPKFIPLKAV--TTGSVPVSMAT 988  
 QY 987 ISALLPAAVAGNYFLSQTIVTNNAQSCIVONLTP-----GPPVPELOHTNRL 1038  
 DB 989 VLIHPQYTKERKNPLMYLTGVOTDKAGDISCANADIMPLKIGOTSSVSFSKSNFHTKXL 1048  
 QY 1039 NSGNTQCVVRCGLGOLAGTEVSGLLALVHNEPFRRAKFSITVVSFFELGTBEGSVL 1098  
 DB 1049 NCRITASCNVTCMLKDVHMKGEYFVAVVTRIMNGTFASSTFQVOLTAAABINTYNPEYI 1108

QY 1099 QLTASRMSSELEVVQTRPILISLMI-----LIGSVGLGLLALLVFCIM 1145  
 DB 1109 -----VIEDNTVTPIMMKPEKAEVPTGVIGSIAGILLALLVALIM 1154  
 QY 1146 KLGFFAKH-----KTFEE 1158  
 DB 1155 KLGFFKRYEKMTKNPDE 1172  
 RESULT 6  
 ITA2\_BOVIN  
 ID ITA2\_BOVIN STANDARD; PRT; 1170 AA.  
 AC P53710;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GP1a)  
 GN (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCB1\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94193647; PubMed=7511592;  
 RA Kamata T., Puzon W., Takada Y.;  
 RT "Identification of putative ligand binding sites within I domain of  
 RT integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";  
 RL J. Biol. Chem. 269:9659-9663(1994).  
 CC -I- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,  
 CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT  
 CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN  
 CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER  
 CC CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE  
 CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
 CC EXTRACELLULAR MATRIX.  
 CC -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
 CC ASSOCIATES WITH BETA-1.  
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -I- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-eb.ch](mailto:license@isb-eb.ch)).  
 CC  
 CC EMBL, L25886; AAB59255.1; -.  
 CC HSSP: P17301; IAOX.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR InterPro: IPR02035; VWF\_A.  
 DR Pfam: PF00092; vwa; 1.  
 DR Pfam: PF00357; Integrin\_A; 1.  
 DR Pfam: PF01839; FG-GAP; 5.  
 DR SMART: SMO0191; Int\_alpha; 5.  
 DR SMART: SMO0327; VWF\_1.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE: PS50234; VWFA; 1.  
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Platelet; Signal; Repeat; Polymorphism; Calcium; Magnesium.  
 FT NON TER 1 1  
 FT SIGNAL 1 18  
 FT CHAIN 19 1170 POTENTIAL.  
 FT DOMAIN 19 1121 INTEGRIN ALPHA-2.  
 FT TRANSMEM 1122 1143 EXTRACELLULAR (POTENTIAL).  
 FT POTENTIAL.

```

FT DOMAIN 1144 1170 CYTOPLASMIC (POTENTIAL).
FT REPEAT 34 92 FG-GAP 1.
FT REPEAT ? 2 FG-GAP 2.
FT DOMAIN 177 367 VMEFA.
FT REPEAT ? 2 FG-GAP 3.
FT REPEAT 423 475 FG-GAP 4.
FT REPEAT 477 538 FG-GAP 5.
FT REPEAT 540 599 FG-GAP 6.
FT REPEAT 604 656 FG-GAP 7.
FT CA_BIND 488 496 POTENTIAL.
FT CA_BIND 552 560 POTENTIAL.
FT CA_BIND 616 624 POTENTIAL.
FT SITE 472 474 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1146 1150 GEFKR MOTIF.
FT DISULFID 72 81 BY SIMILARITY.
FT DISULFID 669 726 BY SIMILARITY.
FT DISULFID 778 784 BY SIMILARITY.
FT DISULFID 854 865 BY SIMILARITY.
FT DISULFID 1008 1039 BY SIMILARITY.
FT CARBOHYD 1044 1049 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1063 1063 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 580 580 G -> V.
FT VARIANT 588 588 R -> K.
FT VARIANT 725 725 R -> S.
SQ SEQUENCE 1170 AA; 128929 MW; EECERIGSF2448FB1 CRC64;

```

Query Match 28.7%; Score 1754; DB 1; Length 1170;  
 Best Local Similarity 33.9%; Pred. No. 2.8e-118;  
 Matches 411; Conservative 235; Mismatches 465; Indels 102; Gaps 30;

```

QY 9 LFLPLVFLTLG---CSFRLDHEHRLPPGPEAFGYSVLOHVGGGGGMVLVGAHPMGF 65
DB 2 LQLVTVFQGLNCCVAVNVGLPKAKIFSGPSSBOFYAVVOQFIPKGMWLVGSPMGF 61
QY 66 SGDRGDVYRCFVGGAHNAFCAGHLL-GDYOLGNSHPVANNHGLMSLLETDDGGFMAC 124
DB 62 PKNRKGDVYKCPV-DLSTTTCKLNLOTSTMSNVTENKTNLSGLILTRNVGTGGLTLC 120
QY 125 APLMSRACGSSVFSGGICARVDASFOPQSLAPTAQRCPTVMDVIVLDGNSIYPMSEV 184
DB 121 GPLMAGQCGSQYTTGVCSDVSPDFQLRTSFAPAVQTCPSFIDVVVVCDESISYPMNAV 180
QY 185 QTFLLRLVGLKFLIDEPQIOVGLVOYGESPHEMSIGDRTKEVVRRAKNLSRRGRGRTK 244
DB 181 KNFLKFPQGLDIGTKTQWGLIOYANNRPVVFNNTKSKDEMKAISOTFOYXGDDTLN 240
QY 245 TQAQITVACTBEGFSGHSGRPRARLLVYVTDGSHDEBELPAALKADEAGVTRGIIAV 304
DB 241 TFKAIOYARDIAYSTAAGRPATKVMVYVTDGSHDEBELPAALKADEAGVTRGIIAV 300
QY 305 LGHYLRORDPSSFLREIRITIASDPERFFNVTDDEALTDVDAIGRIPLGEGSHAHN 364
DB 301 LGYLRNMLDPTKNLKEIKALIASIPTEHFFNVSDADBLEKAKGTIGIGIFIEGT-VQG 359
QY 365 ESSFLEMSQIGFSTRLLKDG--IIFGMVAYDMGGSVLM-LEGHRLFPFMALEDEFP 421
DB 360 GDNFMQEMSQVGFSAEYSPONNIIMLGAVGAYDMGSGTVQKTPHHLIFS-----KQAFE 414
QY 422 PALQ--NHAAYLYGVSVMILGGRPLFSGAPRRHGRKVTAFOLEKKGAVRAVNOISLG 479
DB 415 QILQDRNHSSTYIGIVAS--ISTGNSVHFVAGAPRANVTQIYLVSNENGNATVYIQSORG 473
QY 480 EOIGSYFGSELCPRLDTRDGTDTVLVLAAPMFLGPNKETGRVYVYLVGQOSLTLTGTL 539

```

```

DB 474 DOIISYFSSVLCADVNDKTTITDVLVGAHPVANDLKEBGRVYLFITKILMHQFLE 533
QY 540 QPEPPDARFPGFAMGALPDNLQDGFADVAVAGAPLEDHOGALYHYHGTOSVRRPBPAPRI 599
DB 534 GPNGLNARFGSALIALSDIMMDFNDYVCSPLENONGAVIYVNHGEMIRLRYSOKI 593
QY 600 AAASMPHA--LSYFGSVDRGLDLGDGLVDVAVAGAAIILSSRPVHLTPSLVETPQ 657
DB 594 LGSBRAPFSHLQYFGRSIDGVDLNGDSITDVSVGARGQVYVQLMSQSIADVSVASFPK 653
QY 658 AIVSVQDRCRRRGDAVCLTALCFQVTSRTPGRMDHOFYKRFASLDEMTAG-----ARA 713
DB 654 KITLANKAAE-----IKLKLCFSAKFR-PTNQNNOVAIVNIITDEDEPFSSRISRG 704
QY 714 AFDSGGRLSRRRLRSVGNVTCQQLHFNHVDSTDYLRPVALTPTFALDNTTKG--PVL 771
DB 705 LFKENNEKLOKTIWISQAO-RCEYIITHIDPSDILSPNLKWNISLEN--PQTNPAL 760
QY 772 NEGSPTSIQKLVPFSKDCGPNECVTDLVQVNMDDIGSRAPRVNGRKRVLTSTLE 831
DB 761 EAYSETVAVFSIPIPHKCGDDGVCISDVLVNV-QOLPATQOQPIVSNQNRILTFVQLK 819
QY 832 NRKENAVTSLSTIFSRNHLASTLPQRESPT--KYEC-AAPSAHARLCSVGHVPFQTG 887
DB 820 NKKSAYVTEIVDPFSENLFPASWS---MEVDGTEVYVQIASSQKSVTCNVGYPALKSK 875
QY 888 AKVFFLEFEESGSLISOVGKLTASDSLERGTLOENQAQTSAYQVEPHLLFSES 947
DB 876 QOVFTTFIDFNLOLQWASISFRALSSQSEEN--MADNSVNLKSLIYLAETHIRST 933
QY 948 TLHRYE-----VHPYGLPVGPGPEKTTLRVONLCGYVNSGLIISALLPAVANG 997
DB 934 NINYEVSLLDGNVSVHSPFEDI---GPKTFISIKV-TTGSVYVMSASVYIHIPOYTKD 988
QY 998 GNYFLSLSQVITNNA---SCIVONLTBPPGPVHP-----EELQHTNRING 1040
DB 989 KNPLMYLTVGATDQAGDISCAE-----INPLKIGQTSSSVSFKEENFRHIKEINLC 1039
QY 1041 SNTQCOVVRCHLGLQAKTEVSVGLRLVHNEFFRRAKFKSLYVSTRELTGTESSVQL 1100
DB 1040 RTASCSTNMCWLRDQVKGFEFLVNVSTRIMNGTPAASFPQVQLAAEIDITVNPQIYVI 1099
QY 1101 TEASRWSESLLEVQVTRP---ILISLWILISVGLLILALLVFCMLKGFPAHK----- 1153
DB 1100 EE-----NTVTIPLTIMKHEVVEVPTGVYGSVLAGIILLALVAILMKLGFPRKRYEKM 1155
QY 1154 -KIEEEKREBEKL 1165
DB 1156 AKNPDETDTETEL 1168

```

RESULT 7  
 ITA2\_MOUSE STRAND; PRT; 1178 AA.  
 AC 062469; 062163;  
 DT 15-JUN-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-1998 (Rel. 36, Last sequence update)  
 DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GP1a)  
 GN (Collagen receptor) (VLA-2 alpha chain) (CD49b).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=C57BL/6 X CBA; TISSUE=lung;  
 RX Edelman U.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,  
 RA Danjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;  
 "The mouse VLA-2 homologue supports collagen and laminin adhesion but

RT not virus binding.";  
 RL Cell Adhes. Commun. 2:131-143(1994).  
 RN [2]  
 RP SEQUENCE OF 450-1178 FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=94355691; PubMed=7521231;  
 RA Wu J.E., Santoro S.A.;  
 RT "Complex patterns of expression suggest extensive roles for the alpha  
 RT 2 beta 1 integrin in murine development.";  
 RL Dev. Dyn. 199:292-314(1994).  
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A COLLAGEN RECEPTOR, BEING  
 CC RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER CELLS TO  
 CC COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE EXPRESSION,  
 CC FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
 CC EXTRACELLULAR MATRIX. IT IS ALSO A RECEPTOR FOR LAMININS, COLLAGEN  
 CC C-PROPEPTIDES AND E-CADHERIN. MICE HOMOZOUS FOR A NULL MUTATION  
 CC IN THE ALPHA-2 DIE VERY EARLY IN EMBRYOGENESIS.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VMPA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z29987; CAA82877.1; -;  
 DR EMBL; X75427; CAA53178.1; -;  
 DR HSSP; P17301; IAOX.  
 DR MGD; MGI:96600; Itga2.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VMP\_A.  
 DR Pfam; PF00092; vwa; 1.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR PRINTS; PR00453; VMPADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS50234; VMPA; 1.  
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 DR Platelet; Signal; Repeat; Calcium; Magnesium.  
 KW SIGNAL; 1  
 FT CHAIN 1 1178 INTEGRIN ALPHA-2.  
 FT DOMAIN 27 1129 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1130 1151 POTENTIAL.  
 FT CYTOPLASM 1152 1178 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 42 100 FG-GAP 1.  
 FT REPEAT 185 375 FG-GAP 2.  
 FT REPEAT 185 375 VMPA.  
 FT REPEAT 431 483 FG-GAP 3.  
 FT REPEAT 485 546 FG-GAP 4.  
 FT REPEAT 548 607 FG-GAP 5.  
 FT REPEAT 612 664 FG-GAP 6.  
 FT REPEAT 664 717 FG-GAP 7.  
 FT CA\_BIND 496 504 POTENTIAL.  
 FT CA\_BIND 560 568 POTENTIAL.  
 FT CA\_BIND 624 632 POTENTIAL.  
 FT SITE 480 482 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 1154 1158 GPEPR MOTIF.  
 FT DISULFID 677 89 BY SIMILARITY.  
 FT DISULFID 786 794 BY SIMILARITY.  
 FT DISULFID 862 873 BY SIMILARITY.  
 FT DISULFID 1016 1047 BY SIMILARITY.

FT DISULFID 1052 1057 BY SIMILARITY.  
 FT CARBOHYD 102 102 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 109 109 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 429 429 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 457 457 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 472 472 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 696 696 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1054 1054 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1078 1078 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;  
 Query Match 28.6%; Score 1744; DB 1; Length 1178;  
 Best Local Similarity 35.3%; Pred. No. 1.5e-117;  
 Matches 420; Conservative 221; Mismatches 479; Indels 70; Gaps 31;  
 9 LFLPLVFLT-GL--CSPPNLDHHRRLFPPEAEFGYSVLQHVGGGRMLVCAPMQGP 65  
 10 LLLQLMLLVQGLINCLAVNGLPGAKIFGSGSSSEQFGYSVOQLTPQGNMLLVGSPWGSF 69  
 66 SGDRKGVYRCRPGVGAHNAKCAKGL-GDYOLGNSHFAVMHIGMSLLETDGCGFMAC 124  
 70 PENRGDVYKCPV-DLPATCEKLNLSASISNTEIKTNMSLGLTLRNPDTGGLTC 128  
 125 APLMRACSSVFSGICARVDASFPOGSLAPTACRPTVMVYIVLDGNSIYPMSEV 184  
 129 GPLMWHQGNQYVATGICSDVSPDFQPLTFSPPAQACPSLVDDVYVDEGNSIYPMSEV 188  
 185 QTFELRLVGLKFLPDEQIQVGLVOGESPVHMSIGDFRTKEEVYRAKNLSRREGRETK 244  
 189 KNFLVFKFTGLDIGRKTQVALIQYANPRITFININDETEKEDMWQATSETRHOGGLTN 248  
 245 TQAQIMVACTGSGFSGSHGGRPEARLLVYVVDGSHDEBELPALAKAEAGRVTRYGAV 304  
 249 TFRALFEPADYAYVSGRPGATKVVVVVVDDGSHDSKKTYYIQCNDEILRFGLAV 308  
 305 LGHYLRORDPSSFLREIRTIASDPDERFFNVYDEALTDIVDALGRIFGLEGSHAEN 364  
 309 LGYLNRNALDTRKNLKEIKALIASPTERYFFNVADDEALLERAGTLGQIFSIEST-VQG 367  
 365 ESSFGLENSQIGFSTHRL-KDGLIFGVNAGVADKGSVLMLEGHR--LFFPRMALDEBF 420  
 368 GDNFMENAVQGFSAADYAPONDILMLGAVGAFDWSGLTV-DETSKRPVIFP---KQAF 421  
 421 PPALO--NHAAYVIGSVSMILRGGRFLFLGAPFRFRGVIAFOLKKDGAVRVAGSLQ 478  
 422 DQVLDRNHSSFLGVSVAISTEDGV-FVAGAPRANVTGQVLVSVKKGQNVYIYQSHR 480  
 479 GEOLGSYFGESELCPLDTRDGTDDVLLVAAPMFLGPONKETGRVYVVLVGOQSLTLTQGT 538  
 481 GQOIGSYGVSLCSVDVDKDTITDVLVGAFLPYNMMDLKEGKQVLPITITGILNQHF 540  
 539 LQPEPPDARFGPMGALPDLNQDGFADVAVGAPLEDHQGALYIYHGTQGVRRPQR 598  
 541 EGPEETGARFGSALALASDINMGFNDVIGSPVENNSGAVYVYNHOGTIRKYSQK 600  
 599 IAAA--SMPHALSYGRSGVGRDLDDGDDLVDAVGAQGAAILLSRPIVHLTPLEATP 656  
 601 ILGSGARFRHLPQFGRSLDGYGDNGSDITDVSIGALGOYIQLMSQSLADVAIALTLP 660  
 657 QAISSVQRCRRRGGEAVCLTALCFQVTSRTPGRMHDQFYMRFTASLDEWTAGARAAD 716  
 661 DKITILNNDK-----ITLKLCFRAFFPPAG-NNQVALIFNMTLDADGHSRRVSR 711  
 717 GSGQRLSPRLR--LSGVNV-TCEQLHFLVLDTSYLRPVALTYTFALDNTTKP--FVL 771  
 712 GVFRNSERFLQKMNVAIVEQKSEHHISIQKPSDVVNPDLDRYDISLEN--PQTSPAL 768  
 772 NEGSTSIQKLVPPSKDQDNECTDVLQVNMDIRSKRAPFVVRGGRKVLVSTLE 831  
 769 EAYSTVAVFSIPFYKCGSDGICISDLITLV-QQLPAIQOSFVSNQRRLTFSVLK 827  
 832 NRKENAVNTSLSIFSRMLHSLTPQRESPI---KYECAPSAHARL-CSVGHPVFQTC 887

```

Db      828 NGESAYNTVVAEFAEENLEFASFS-----MPVDGTVTEVGSQKSVTCVGYPALKSE 883
Oy      888 AKVTFLEEFSSCSLSQVFGKLTASPSLENGTLOENTQTSAYIQEPHLLFSSSS 947
      884 QCVTFINFDENQONQONQAINFOAFSSQETNKA--DNSVSLTPLIYDELHLTRST 941
Oy      948 THREYVHPYGLP-----VGPPEFKTLRVQNLGCYVSGLLISALLPAVAGGNVF 1001
      942 NINFEYSSDENAPSVYKSVEDIGPKFISLKV--TAGSAPVSMALVTIHIPQYTEKNPDL 1000
Oy      1002 LLSQVITNNA---SCL--VQNLTEP---PGPPVHEELQHTNRNLNGSTQCQVVRCHLG 1053
      1001 LVLGTQTDAGDISCTAEINPKLPHTPASVSFKXNENRHTKELDCRTTSCSNITCWLK 1060
Oy      1054 QAKGTEVSVGLRLVHNEFFRRAKFSLTAVSTFELGTEEGSVLQLTASRMSLSLELV 1113
      1061 DHMAKEVINTVTRWNRTEFAASTQVQLTAAEIDITHNQ-LFVIEENAVTIPLMIM 1119
Oy      1114 VQTPRLISLWILIGSVLGLLLALLVFCWLKGF--AHKKI---PEE 1158
      1120 KPEKAEVPTVGIISIGIILLMTAGLWKLGFGRKQYKMGQNPDE 1169

```

## RESULT 8

```

ID      ITAD_HUMAN      STANDARD;      PRT; 1162 AA.
AC      Q13349; Q15575; Q15576;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DE      15-OCT-2002 (Rel. 41, Last annotation update)
DE      Integrin alpha-D precursor (Leukointegrin alpha D) (CD11d) (ADB2).
GN      ITGAD.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Spleen;
RX      MEDLINE=96111956; PubMed=8777714;
      Van der Vlieten M., Le Trong H., Wood C.L., Moore P.F., St John T.,
      Staunton D.E., Gallatin W.M.;
      "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-
      3.";
      Immunity 3:683-690(1995).
RN      [2]
RP      SEQUENCE OF 1-235 FROM N.A.
RX      MEDLINE=20187620; PubMed=10722744;
      Nott J.D., Johnson A.K., Dillon J.D.;
      "Structural and functional characterization of the leukocyte integrin
      gene CD11d. Essential role of Spl and Sp3.";
      J. Biol. Chem. 275:8959-8965(2000).
RN      [3]
RP      SEQUENCE OF 467-537; 571-602; 633-666; 788-834 AND 910-1125 FROM N.A.
RX      MEDLINE=96257236; PubMed=8666289;
      Wong P.A., Davis B.M., LeBeau M., Springer T.A.;
      "Cloning and chromosomal localization of a novel gene-encoding a human
      beta 2-integrin alpha subunit.";
      Gene 171:291-294(1996).
RN      [4]
RP      INTERACTION WITH VCAM1.
RX      MEDLINE=99059842; PubMed=9841932;
      Grayson M.H., Van der Vlieten M., Sterbinsky S.A., Michael Gallatin W.,
      Hoffman P.A., Staunton D.E., Bochner B.S.;
      "alpha2beta2 integrin is expressed on human eosinophils and functions
      as an alternative ligand for vascular cell adhesion molecule 1
      (VCAM-1).";
      J. Exp. Med. 188:2187-2191(1998).
RN      [5]
RP      INTERACTION WITH VCAM1.
RX      MEDLINE=99370002; PubMed=10438935;
      Van der Vlieten M., Crowe D.T., Hoekstra D., Vazeux R., Hoffman P.A.,

```

```

RA      Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;
RT      "The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a
      binding interface between I domain and VCAM-1.";
      J. Immunol. 163:1984-1990(1999).
RL      [1]
CC      FUNCTION: INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM1 AND
      VCAM1. MAY PLAY A ROLE IN THE ATHEROSCLEROTIC PROCESS SUCH AS
      CLEARING LIPOPROTEINS FROM PLAQUES AND IN PHAGOCYTOSIS OF BLOOD-
      BORNE PATHOGENS, PARTICULATE MATTER, AND SENESCENT ERYTHROCYTES
      FROM THE BLOOD.
CC      SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-D
      ASSOCIATES WITH BETA-2.
CC      SUBCELLULAR LOCATION: Type I membrane protein.
CC      TISSUE SPECIFICITY: EXPRESSED MODERATELY ON MELANOCYTIC CELL
      LINES AND SUBSETS OF PERIPHERAL BLOOD LEUKOCYTES AND STRONGLY ON
      TISSUE-SPECIALIZED CELLS, INCLUDING MACROPHAGES FOAM CELLS WITHIN
      ATHEROSCLEROTIC PLAQUES, AND ON SPLENIC RED PULP MACROPHAGES.
CC      DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WFPA DOMAIN. INTEGRINS
      WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC      SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC      SIMILARITY: CONTAINS 1 WFPA DOMAIN.
CC      SIMILARITY: CONTAINS 7 FG-GAP-REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
      the European Bioinformatics Institute. There are no restrictions on its
      use by non-profit institutions as long as its content is in no way
      modified and this statement is not removed. Usage by and for commercial
      entities requires a license agreement (see http://www.isb-sib.ch/announce/
      or send an email to license@isb-sib.ch).
      -----
DR      EMBL; U37028; AAB38547.1; -
DR      EMBL; U40274; AAB60634.1; -
DR      EMBL; U40275; AAB60635.1; -
DR      EMBL; U40276; AAB60636.1; -
DR      EMBL; U40277; AAB60637.1; -
DR      EMBL; U40279; AAB60638.1; -
DR      EMBL; U40278; AAB60638.1; JOINED.
DR      HSSP; AF187881; AAF62875.1; -
DR      HSSP; P11215; IABX.
DR      Genew; HGNC:6146; ITGAD.
DR      MIM; 602453; -
DR      InterPro; IPR000413; Integrin_alpha.
DR      InterPro; IPR002035; WFPA.
DR      Pfam; PF00092; vwa; 1.
DR      Pfam; PF00357; Integrin_A; 1.
DR      Pfam; PF01839; FG-GAP_5.
DR      PRINTS; PR01185; INTEGRINA.
DR      PRINTS; PR00453; WFPADOMAIN.
DR      SMART; SM00191; Int_alpha; 4.
DR      SMART; SM00327; VMA; 1.
DR      PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR      PROSITE; PS50234; WMPA; 1.
KW      Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW      Signal; Repeat; Calcium; Magnesium.
FT      SIGNAL          1      17
FT      CHAIN           18     1162
FT      DOMAIN          1101    1124
FT      TRANSMEM        1125    1162
FT      DOMAIN          1125     85
FT      REPEAT          32      85
FT      REPEAT          150     332
FT      REPEAT          350     400
FT      REPEAT          401     452
FT      REPEAT          454     516
FT      REPEAT          518     576
FT      REPEAT          581     633
FT      CA_BIND         465     473
FT      CA_BIND         530     538
FT      CA_BIND         593     601
FT      SITE            1127    1131
FT      DISULFID        67      74
FT      POTENTIAL.
FT      INTEGRIN ALPHA-D.
FT      EXTRACELLULAR (POTENTIAL).
FT      POTENTIAL.
FT      CYTOPLASMIC (POTENTIAL).
FT      FG-GAP 1.
FT      FG-GAP 2.
FT      WMPA.
FT      FG-GAP 3.
FT      FG-GAP 4.
FT      FG-GAP 5.
FT      FG-GAP 6.
FT      FG-GAP 7.
FT      POTENTIAL.
FT      POTENTIAL.
FT      POTENTIAL.
FT      GFPR MOTIF.
FT      BY SIMILARITY.

```

```

FT DISULFID 106 124 BY SIMILARITY.
FT DISULFID 655 710 BY SIMILARITY.
FT DISULFID 769 775 BY SIMILARITY.
FT DISULFID 846 861 BY SIMILARITY.
FT DISULFID 994 1018 BY SIMILARITY.
FT DISULFID 1023 1028 BY SIMILARITY.
FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 733 733 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 873 873 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 500 500 MISSING (IN REF. 2).
FT CONFLICT 515 518 GHW -> ATP (IN REF. 2).
FT CONFLICT 825 825 L -> V (IN REF. 2).
FT CONFLICT 984 984 V -> A (IN REF. 2).
SQ SEQUENCE 1162 AA; 126885 MM; F296A1A35455D77D CRC64;

```

Query Match 18.5%; Score 1131.5; DB 1; Length 1162;  
 Best Local Similarity 29.4%; Pred. No. 2,3e-73;  
 Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

```

23 FNLDENHRLPFGPPEAEFGYSVLQVHGQGMVLVGAPEWDPGSGDRGDVYRCVGAH 82
18 FNLDVEEPIIFQ-EDAGGFGQSVQF--GGR-LVVGAPLEVAANOTGRIDC---AAA 70
83 NAPCAKGLGDYQLGNSSHPAVNMHLSLETGQGGFMACAPLMSRACGSSVFSSGIC 142
71 TGMCGPIPL-----HIREAVNMSLGLTLASTNGSRLACGPTLHRVGENSEYKSGC 124
143 ARVDASFOGQSLAPFAQRCP--TYMDVVIVLQGSNFI--YPMSEVQTFRLRLVKGFI 199
125 LILGSMWEITQVDPATPECPHOEMDIVFLIDSSGSIIDONDFQMGMFOQAVWQ--FEG 182
200 EOIQVGLVOYGESPVHEMSLGPRTKEEVYRAKNISSREGRETKQAQIMVACTEGFSQ 259
183 TDTLPMQYSNLKHFTFTQRTSPSQSLVDPVQLKL-TFRTATGILVTVLDFH 241
260 SHGGRPEARLLVVTGESH--DGEELPALKACEAGRTYRGIAVLGHYLRQRDPSSF 318
242 KNGARSAKKILIVTDGQKYKDPLEYSDVIPAOKAGIIRYAGV-GHAF---QGPTA- 296
319 LBEIRIADDPDRFFPNVTDEALDIYDALGDRIFGLEGAENESSFGLMSGIGTS 378
297 ROELNTISSAPPQDHVFKVDNFPALGSIQKQLOEKIYAVEGTQSRASSSFQHEMSQEGFS 356
379 TRLKDGILFGWVGAVDMGGSVLMLEGHRLFPFRAALDEDEFPALQNH-----AY 430
357 TALTMGLFLGAVGSFSW-----SGNAFLYPPNMS-----PTINMSQENVDRDY 403
431 LGYSSSMLLRGRRFLSGAPFRRRKGVIAF-QLKDGAVRVAOSLQEOIGSYFSGE 489
404 LGYSTELALMKGVQNLV--GAPRYQHTGKAVIFTYQSRQ--RKKALEVITGTQGSYFGAS 460
490 LCPDLDROGTTDVLLVAPMFLGPNKERTGYVVV--GQSSLLTLOGTLQPEPPD- 546
461 LGSVDVDSGSTDLILIGAPHY--KOTRGQVSVCPPLPRGQGVQWQCAVALRGEQHPW 518
547 ARFGFAGALPDLNODGFADVAVGAPLEDGOGALVLYHG-TQSGVRPPHQAQIAASMP 605
519 GRFGALATVLDGVDNEKLDVAIGAGCEQENRCAVILFHGASGSGISPHSQRILASSQS 578
606 HALSYFGRSVDGLDGDLDVAVAGQAAILSSRPVHLTPSLEVYTPQAISVQRD 665
579 PRLOYGQALSGGQDLOGMLDLAVAGAGQVLLRSLPVLKVGAMRSPSEVAAVVR 638
666 C-----RRGQAVCTIATLCTQVTSRTPGKWDHQFMRFPASIDENVTAGARAFDS 718
639 CMEKPSALEAGATVCTLTQ-----KSSLDQIGDIQSSVRFDLADPRLTSLRAIFNET 693

```

```

QY 719 GQRLSPRLRLSVGNVTCEQHLFVLD-TSDYLRVALVTVPALDNTKPG-----PVLN 772
DB 694 KNPULTRKRTIGLG-ICETELKLLPDCVEDVSPILIHNSLVRBEPISPQNLRPVLA 752
QY 773 EGSPTSIQKLVFSPSCDCEPNECVTDVLQVNMIDRSRKAPFVVRGGRKKLVSTTLEN 832
DB 753 VGSQDLFTASIPFEKNCCODGICEBD--LGVTLSFSGIQT--LVVGSLSLELNVITVWN 807
QY 833 KRENAYNTSLGIIIFARNLHSLATPQRES-----IKVBC-AAPSAHARL---CSVGHV 883
DB 808 AGEDSVYGVSLYVAGLSHRVSGAQKOPHOSALRLACEYFTPEDEGLRSRCGVNHI 867
QY 884 FQTKAVTFLLEFEPSCSLLSQVFGKLTASSDSLERNGLQENTAGTSAYIQEPHLLF 943
DB 868 FHGSGNGFIETFDVSYATILGDRM-LMRASASSNNKASSKATQLELPKVAVYVMI 926
QY 944 SEESTLHREYHPYITLVGQEPPEKTLRVQNLGCIYVSGILISALLPAVHAGNYFLS 1003
DB 927 SROESTRY--FNFAISDEKKKKEAHRVNNLSQBDLA-ISLFWFVLLNGVAVADV 983
QY 1004 LQVITNNASCIVQNLTEPPGPVPAPEELQHTNR---LNGSNVQGVVRCGLQAKGTE 1060
DB 984 VNEAPSGSLPCVSEK-----KPPQHSDFLTQISRPMLDCSIADCLQRCQDVPSVQBE 1038
QY 1061 VSVGLRLRVNHEFFRAKFKSLTVVSTFELTEGSSVLQLTASRSESLLE-VVQTRPI 1119
DB 1039 LDFTLKGNLSFGWVARETLQKKVLVVSVAEITFDTSVQSLPQQAFAQMEVMELEDEV 1098
QY 1120 LISLWILIGSVLGLLALLVFCWKLGFF-AHKKIPEEKREE 1163
DB 1099 VNAIPITINGSSVGLALLLALITATLYKLGFRRHYKEMLEKPED 1143

```

## RESULT 9

```

ID ITL_HUMAN STANDARD; PRT; 1170 AA.
AC P20701; Q43746;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-L precursor (Leukocyte adhesion glycoprotein LFA-1
DE alpha chain) (Leukocyte function associated molecule 1, alpha chain)
DE (CD11a).
GN ITGAL OR CD11A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.
RX MEDLINE=89139587; PubMed=2537322;
RA Larson R.S., Corbi A.L., Berman L., Springer T.;
RT "Primary structure of the leukocyte function-associated molecule-1
RT alpha subunit: an integrin with an embedded domain defining a protein
RT superfamily.";
RL J. Cell Biol. 108:703-712(1989).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.U., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrman J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.B., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q.";
RL Genomics 60:295-308(1999).
RN (3)
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 153-335, AND REVISION TO 214.
RX MEDLINE=96036067; PubMed=7479767;
RA Ou A., Leahy D.J.;
RT "Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, alpha
RT L beta 2) integrin.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:10277-10281(1995).

```

[4]  
 X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 153-335.  
 MEDLINE=96398682; PubMed=805579;  
 Qu A., Leahy D.J.;  
 "The role of the divalent cation in the structure of the I domain  
 from the CD11a/CD18 integrin";  
 Structure 4:931-942 (1996).  
 [5]  
 X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 153-334.  
 MEDLINE=99425288; PubMed=10493852;  
 Kallen J., Welzenbach K., Ramage P., Geyl D., Krawacki R., Legge G.,  
 Cottens S., Weitz-Schmidt G., Hommel U.;  
 "Structural basis for LFA-1 inhibition upon Iovastatin binding to the  
 CD11a I-domain";  
 J. Mol. Biol. 292:1-9 (1999).  
 -FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,  
 ICAM3 AND ICAM4. IT IS INVOLVED IN A VARIETY OF IMMUNE PHENOMENA,  
 INCLUDING LEUKOCYTE-ENDOTHELIAL CELL INTERACTION, CYTOTOXIC T-CELL  
 MEDIATED KILLING, AND ANTIBODY DEPENDENT KILLING BY GRANULOCYTES  
 AND MONOCYTES.  
 -SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-L  
 ASSOCIATES WITH BETA-2.  
 -SUBCELLULAR LOCATION: Type I membrane protein.  
 -ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
 produced by alternative splicing.  
 -TISSUE SPECIFICITY: LEUKOCYTES.  
 -DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VMFA DOMAIN. INTEGRINS  
 WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 -SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 -SIMILARITY: CONTAINS 1 VMFA DOMAIN.  
 -SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 -DATABASE: NAME=PROW; NOTE=CD guide CD1a entry;  
 WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd1a.htm".  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL, Y00796; AAC68747.1; -  
 DR EMBL, AC002310; AAC31672.1; -  
 DR PIR, S03308; S03308.  
 DR PDB, 1LEA; 29-JAN-96.  
 DR PDB, 1ZON; 07-DEC-96.  
 DR PDB, 1ZOO; 07-DEC-96.  
 DR PDB, 1ZOP; 07-DEC-96.  
 DR PDB, 1COP; 07-AUG-00.  
 DR Genew; HGNC:6148; ITGAL.  
 DR MIM; 153370; -  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00092; vwa; 1.  
 DR Pfam; PF00357; integrin\_A; 1.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR PRINTS; PR01185; INTEGRINA.  
 DR PRINTS; PR00453; VMFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; vwa; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS00234; VWFA; 1.  
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal; 3d-structure; Magnesium; Calcium; Repeat;  
 KM Alternative splicing.  
 FT CHAIN 1 25  
 FT DOMAIN 26 1170 INTEGRIN ALPHA-L.  
 FT TRANSMEM 1089 1112 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 1113 1170 POTENTIAL.  
 FT REPEAT 42 91 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 92 149 FG-GAP 1.  
 FT REPEAT 92 149 FG-GAP 2.

FT	DOMAIN	170	349	VMFA.
FT	REPEAT	?	?	FG-GAP 3.
FT	REPEAT	401	455	FG-GAP 4.
FT	REPEAT	457	516	FG-GAP 5.
FT	REPEAT	518	575	FG-GAP 6.
FT	REPEAT	578	630	FG-GAP 7.
FT	CA_BIND	468	476	POTENTIAL.
FT	CA_BIND	530	538	POTENTIAL.
FT	CA_BIND	590	598	POTENTIAL.
FT	SITE	1115	1119	GEPR MOTIF.
FT	DISULFID	73	80	BY SIMILARITY.
FT	DISULFID	111	129	BY SIMILARITY.
FT	DISULFID	653	707	BY SIMILARITY.
FT	DISULFID	771	777	BY SIMILARITY.
FT	DISULFID	845	861	BY SIMILARITY.
FT	DISULFID	998	1013	BY SIMILARITY.
FT	DISULFID	1021	1052	BY SIMILARITY.
FT	CARBOHYD	65	65	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	89	89	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	188	188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	649	649	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	670	670	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	726	726	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	730	730	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	862	862	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	885	885	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	897	897	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1060	1060	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1071	1071	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	954	954	Q-> QGVGLVETQTSKQICRPGADMEHTVGAQEGELPC PMGSEAFRDINRAGPCR (IN ISOFORM 2). R-> W (IN REF. 1 AND 2). Y-> I (IN REF. 2).
FT	CONFLICT	214	214	
FT	CONFLICT	660	660	
FT	SEQUENCE	1170 AA; 128819 MW; 39A7AF2EF286FC0 CRC64;		

Query Match 18.1%; Score 1105; DB 1; Length 1170;  
 Best Local Similarity 29.4%; Pred. No. 1.9e-71;  
 Matches 366; Conservative 198; Mismatches 465; Indels 216; Gaps 55;

QY	11	LPVLVLTG-----CSPNLDHHRLLFPGRPEA--EFGSYLVQHVGGQRMVLWGAWMD 63
DB	9	MAMALLSGFFPPAPASSTNLDVRGARSF--SPRARHFGIRVLQ--VVGNG--VVGAGGE 63
QY	64	GPSGDRRDVYRCIPGAGAHNAPCAKGLGDYLGNSHPAVNMILGMSLLETGDCGGEPA 123
DB	64	GNS---TSLVQCQCGSTGHCLPVT-----LRGSNY--TSXYLGMLTADPTDGSILA 110
QY	124	CAPLWSRAAGSSVFSSGICARVDASFQ--PQGSIAFTAGRC--PTYWDVVIYLDGNSNIYP- 180
DB	111	CDPGLSRFCQDQNTVYLSGLCYLFRQNLQGPMLQGRPGQECIKGNVDLVFLFDGSMQLPD 170
QY	181	MSEVQTFLRLRVGLFLFDPEQIQVGLVQYGESPVHMSLDF---RKKEVVRARAKULS 236
DB	171	EFQKILDFMKQVMKKL--SNTSYQPAAVOFSTSYKTEPDSYVVRKPDALDKVKKML 228
QY	237	RREGRETKTAQAINVACTEGFSQSHGRPEARLLVWTDSESHGDELPAALKACEAGR 296
DB	229	LL---TTFGALINVVALEVRREELGARDPKTVLIIITDGEATDSGNIDAAKD----- 278
QY	297	VTRYGIAVLGYLRRQDPSSFLREIRTIASDPDERFFPNVTDEALTDIYDALCDRIFG 356
DB	279	IIRYIIIGIKGFQTESOET-----LHKFAKSPASEEFKILDTFEKLKDLFTELQKIYV 333
QY	357	LEGSHAENSSFGLEMSIGFSTHRLKXGILLFGWGAADMGGSVLMLEGGHRLPPRNAL 416
DB	334	IEGTSKQDTSFNMELSSGSIADISRGHAAGAAGAGGAFDLDAADQ----- 385
QY	417	EDFEF---PPALQNHAAVLGYSVSSMLRGGRRLLFLSGAPFRHKGKVIATQKKDGA-V 471
DB	386	DDTFIGNPLRPEVAVAGLYTVWLPGRQKTSLSASGAPRYQHGKRVLLQDEPGGGHW 445
QY	472	RYAQSLQGEQIGSYFGSELCPDLDTRDGTTVLLVAADMPILGPONKETGRVYVY--LVG 528







```

FT DISULFID 1017 1048 BY SIMILARITY.
FT CARBOHYD 86 185 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 668 668 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 724 724 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 890 890 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 927 927 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1056 1056 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 1163 AA: 128343 MW: A7A3078489BE232F CRC64;

Query Match 18.0%; Score 1097; DB 1; Length 1163;
Best Local Similarity 28.4%; Pred. No. 7.2e-71;
Matches 356; Conservative 192; Mismatches 470; Indels 234; Gaps 48;

12 PLVFLTGL-----CSFNLDEHNRLLFRRPREREGYSTLQHVGGQRMVLGAPMDGS 66
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8 PRLLLGLQLFAKAMSYNLDTRPTQSFLOAGRHFQYQLIEDG---VVVGARPE--- 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GDRRGDVRCFPGAGAHNAPCAKHLADYQLGNSSHPAVMHGLMSLLETDDGGFMACAP 126
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 GNTGGVHCRITSSERCQVS-----LHGSNH--TSKYIGMTLATDAAGSLIADCP 110
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 LMSRAQSSVFSGGICARVDASFQPGSLAPTQRCPTV-----MDVVIYLDGSNSI- 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 GLSRCTQDQNTYLSGLC-----YLPGSLGPMQLQNPAYQCMKQKVDLPLFDGQSID 165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 -YPMSEVQFELRLVYKGLFIDREQIQVGVYQGESVHEHMSLQDPRTEKVEVRAKNLSR 237
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 RQDFEKLIEFMKDVNKRLL--SNTSYQFAAVQSTDRCTEFTFLDY-----VKQNNPQV 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 REGRE-----TKTAQAIWVACTEGFSQSHGRPEAKRLVVVVTGESHDELPALAKA 291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 LIGSVQPMFLNTRAIYVVAHVFKESGARPDATKVLVITTDGASDKGNISA----- 273
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 CEAGVTRVGIYGLVHYLRQRDPSSFLREIRTIASDPEREFENVTDAAITDIDALG 351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 --AHDIRYIIIGIKHFVSVQKO-----KTLHFASEPEVEEFVKLIDPEKIKDLFTDQ 326
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 DRIFGLEGSHAESEFGLMEQIGSTRKLDGILFGVAVYDMGGSVLMLEGHRLLP 411
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 RRIYAIEGTNRQDLTISFNNELSSGISADLSKHAVVAVGAKDMAAGFLDLR----- 379
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 PMALED-----EFPALQNHAAVLGYSVSMLLRGRLPLSGARFRRHGRVLA 463
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
380 -----EDLOGATFVGOEPLTSDVRGGYLYTVAWMTSRSLRLAAGARRVHVGVLVF 434
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 QI-KKDGAIVVAQSLQEOIGSYFSGELCPRLTDRDITDVLVLAAPMLGPONKETGRV 522
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 QAPRAGGRNMQOKIEGTQIGSYFSGELCSVDLDGGEALLIGARLPFGSQ--RGGRV 492
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
523 YVYLVGQGSLLTQGLTLPPEPPD--ARFGFAMGALPDLNQDFADVAVAGARLEDHGQAL 581
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 FTY-QRQSLFPMVSELQDPGVPLQRFMAITALLDINGDLITDVAAGARLEE--QGAV 549
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
582 YLYHGQSGVRRPAPQRIAAASMPHALSYFGRSVSDGRLLDDGDLVDVAVAGQAAILIS 641
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
550 YIFNGKPGGLSPQPSQRIQGAOVFPGIRWFGRSIHVVKDLGGRLADVAVAGGRVVLIS 609
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
642 SRPIVHLTSLLEVTPQAIISVQVQDC-----RRGOEAVCTLAALCFQVTSRTFGRMDFQY 697
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 SRPVVDVIVELSSPEEIRPVHEVEGYSAREQKGVKKA--CFRIKVLTP-----QIQ 662
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
698 MR-----FTASIDETAGARAAFDGSGGRSLFRRLRLSVGV-----TCQQLHNV-L 744
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 GLRLANLSTYLLQDGHMRMRSGLLFPDGSHEL-----GNTSITPDKSCUDPFHHPFI 714

```

```

QY 745 DTSYLREPALVTVPALDNTTKPGPVINEGSP-----SIQKLVPSK 788
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 745 CIQDILSPINSLNLSL-----LEBEGTPROCKGRAMQILRPSIHTVTKEIFEKN 766
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 789 CGPDNECTDLVQVNMIDIRSKRAPFVVRGGRKVLVSTLLENKEMAYNTSLIISR 848
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 767 CGEDKCCBANLTL-----SPARSGPRLIMSS--ASLAEVTLSSNGEDAVWVRDLDP 820
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 849 NLHLASLTP-QRESPIKVECAAPSAHARL-----CSGVHPVPQTGAQVTFLEPEFSCS 901
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 821 GLSPKAVMLQPHSMRPMPSCEELTEGSSLTKTKCNVSPITFAGQGV----- 870
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 902 SLISQVFKLTASS--DSLERNGLT-----QENTAQTSAYIQEPIHLSSSESTL 949
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 871 --LQWFMFTLNSGMEDEVELNGVHCENENSSIQEDNSAATHIPVLPVAILTFEQGN- 927
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 950 HRYEVHPIGTIPVQGPPEFKTKLKYQNL-----GCYVSGGLIISALLPAVHAGNIFL 1002
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 928 -----STLYISPTPGPKTQOVQVHYOVRIOPSADYHMPLEALVGVPRPSEDLI 979
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1003 SLISQVITNN--ASCIQVNLTEPPGPPVPRPELQHTNRNGSNTQCOVY--RCHLQGLAKGT 1059
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 960 TYTMSVQTDPLVYCHSEDLKRPSSBAEP-----CLPGVQFCPIVFRWEILIOVTGT 1032
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1060 EVSVGLRLVHNEPFRRAKFKS--LTVVSTFELGTEEGSVLQLTENSRWSES-----LREVV 1114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1033 V-----ELSKETKASTLSLGSLSVFSNFGHFL--YGSKASEAVQVVKVDLI 1080
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1115 QTRPILSLMILIGSVLGLLILALVFLCKLKGFPFAHKIPEEKREEKLE 1166
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1081 HEKEML-HVUYLSG--IGLVLLFLFLALYKVGFF-----KRNLEKME 1122
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 11

```

ID ITAM HUMAN STANDARD; PRT; 1152 AA.
AC P11215;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha
DE subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MO1)
DE (Neutrophil adherence receptor).
GN ITGAM OR CR3A OR CD11B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86315033; Pubmed=2457584;
RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;
RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor
RT type 3, CD11b) alpha subunit. Cloning, primary structure, and
RT relation to the integrins, von Willebrand factor and factor B.";
RL J. Biol. Chem. 263:12403-12411 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88190151; Pubmed=2833753;
RA Arnout M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;
RT "Molecular cloning of the alpha subunit of human and guinea pig
RT leukocyte adhesion glycoprotein MO1: chromosomal localization and
RT homology to the alpha subunits of integrins.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780 (1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=88257215; Pubmed=2454931;
RA Arnout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;
RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion
RT receptor MO1 (complement receptor type 3).";
RL J. Cell Biol. 106:2153-2158 (1988).
RN [4]

```

RP SEQUENCE FROM N.A.  
RA MEDLINE=93123748; PubMed=8419480;  
RX Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.F., Tenen D.G.;  
RT "Structural analysis of the CD11b gene and phylogenetic analysis of  
RT the alpha-integrin gene family demonstrate remarkable conservation of  
RT genomic organization and suggest early diversification during  
RT evolution.";  
RL J. Immunol. 150:480-490(1993).  
RN [15]  
RP SEQUENCE OF 9-1153 FROM N.A.  
RX MEDLINE=89098893; PubMed=2563162;  
RA Hickey D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L.,  
RA Roth G.J.;  
RT "cdna sequence for the alpha M subunit of the human neutrophil  
RT adherence receptor indicates homology to integrin alpha subunits.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).  
RN [16]  
RP SEQUENCE OF 1-9 FROM N.A.  
RX MEDLINE=92073318; PubMed=1683702;  
RA Shelley C.S., Arnaut M.A.;  
RT "The promoter of the CD11b gene directs myeloid-specific and  
RT developmentally regulated expression";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).  
RN [17]  
RP SEQUENCE OF 1-9 FROM N.A.  
RX TISSUE=Blood;  
RA MEDLINE=92144986; PubMed=1346576;  
RA Pahl H.L., Rosmarin A.G., Tenen D.G.;  
RT "Characterization of the myeloid-specific CD11b promoter.";  
RL Blood 79:865-870(1992).  
RN [18]  
RP SEQUENCE OF 17-31.  
RX MEDLINE=87076671; PubMed=3539202;  
RA Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnaut M.A.;  
RT "N-terminal sequence of human leukocyte glycoprotein Mol. 1;  
RT conservation across species and homology to platelet Iib/Iiia.";  
RL Biochim. Biophys. Acta 874:368-371(1986).  
RN [19]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.  
RX MEDLINE=95171458; PubMed=7867070;  
RA Lee J.O., Rieu P., Arnaut M.A., Liddington R.;  
RT "Crystal structure of the A domain from the alpha subunit of integrin  
RT CR3 (CD11b/CD18)." ;  
RL Cell 80:631-638(1995).  
RN [110]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
RX MEDLINE=96363671; PubMed=8747460;  
RA Lee J.O., Bankston L.A., Arnaut M.A., Liddington R.C.;  
RT "Two conformations of the integrin A-domain (I-domain): a pathway for  
RT activation?";  
RL Structure 3:1333-1340(1995).  
RN [111]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
RX MEDLINE=98362595; PubMed=9687375;  
RA Baldwin E.T., Sawyer R.W., Bryant G.L. Jr., Curry K.A.,  
RA Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrichson R.L.,  
RA Horton N.C., Kelley L.L., Mildner A.M., Moon J.B., Mott J.E.,  
RA Mutchler V.T., Tomich C.S., Watzenpugh K.D., Wiley V.H.;  
RT "Cation binding to the integrin CD11b I domain and activation model  
RT assessment." ;  
RL Structure 6:923-935(1998).  
RN [112]  
RP 3D-STRUCTURE MODELING OF 17-616.  
RX MEDLINE=98226734; PubMed=9560195;  
RA Oxvig C., Springer T.A.;  
RT "Experimental support for a beta-propeller domain in integrin alpha-  
RT subunits and a calcium binding site on its lower surface." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).  
CC -1- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS  
CC ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES  
CC AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.  
CC IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF  
CC THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D

CC PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR  
CC FIRINGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES  
CC OF FIRINGEN GAMMA CHAIN.  
CC -1- SUBUNIT: HETERO DIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-M  
CC ASSOCIATES WITH BETA-2.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
CC GRANULOCYTES.  
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS  
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.  
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD11b entry;  
CC WWW=<http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm>.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
CC -----  
CC EMBL: J03925; AAA59544.1; -;  
CC EMBL: M18044; AAA59491.1; -;  
CC EMBL: J04145; AAA59903.1; -;  
CC EMBL: S52227; AAB24821.1; -;  
CC EMBL: S52152; AAB24821.1; JOINED.  
CC EMBL: S52153; AAB24821.1; JOINED.  
CC EMBL: S52154; AAB24821.1; JOINED.  
CC EMBL: S52155; AAB24821.1; JOINED.  
CC EMBL: S52157; AAB24821.1; JOINED.  
CC EMBL: S52159; AAB24821.1; JOINED.  
CC EMBL: S52161; AAB24821.1; JOINED.  
CC EMBL: S52164; AAB24821.1; JOINED.  
CC EMBL: S52165; AAB24821.1; JOINED.  
CC EMBL: S52167; AAB24821.1; JOINED.  
CC EMBL: S52169; AAB24821.1; JOINED.  
CC EMBL: S52170; AAB24821.1; JOINED.  
CC EMBL: S52173; AAB24821.1; JOINED.  
CC EMBL: S52174; AAB24821.1; JOINED.  
CC EMBL: S52180; AAB24821.1; JOINED.  
CC EMBL: S52181; AAB24821.1; JOINED.  
CC EMBL: S52184; AAB24821.1; JOINED.  
CC EMBL: S52189; AAB24821.1; JOINED.  
CC EMBL: S52191; AAB24821.1; JOINED.  
CC EMBL: S52192; AAB24821.1; JOINED.  
CC EMBL: S52203; AAB24821.1; JOINED.  
CC EMBL: S52212; AAB24821.1; JOINED.  
CC EMBL: S52213; AAB24821.1; JOINED.  
CC EMBL: S52216; AAB24821.1; JOINED.  
CC EMBL: S52219; AAB24821.1; JOINED.  
CC EMBL: S52220; AAB24821.1; JOINED.  
CC EMBL: S52221; AAB24821.1; JOINED.  
CC EMBL: S52222; AAB24821.1; JOINED.  
CC EMBL: S52226; AAB24821.1; JOINED.  
CC EMBL: M76724; AAA58410.1; -;  
CC EMBL: M84477; AAA51960.1; -;  
CC PIR: A31108; RWHLB.  
CC PIR: A26091; A26091.  
CC PDB: 1A8X; 17-JUN-98.  
CC PDB: 1BHO; 18-NOV-98.  
CC PDB: 1BHO; 18-NOV-98.  
CC PDB: 1IDN; 25-NOV-98.  
CC PDB: 1IDO; 01-AUG-96.  
CC PDB: 1JLM; 11-JAN-97.  
CC Genew; HGNC:6149; ITCAM.  
CC MIM: 120980; -;  
CC InterPro; IPR000413; Integrin\_alpha.  
CC InterPro; IPR002035; VFMA.  
CC Pfam; PF00092; vwa; 1.  
CC Pfam; PF00357; Integrin\_A; 1.



CC - FUNCTION: INTEGRIN ALPHA-X/BETA-2 IS A RECEPTOR FOR FIBRINOGEN. IT  
 CC RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN. IT MEDIATES CELL-CELL  
 CC INTERACTION DURING INFLAMMATORY RESPONSES. IT IS ESPECIALLY  
 CC IMPORTANT IN MONOCYTE ADHESION AND CHEMOTAXIS.  
 CC - SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-X  
 CC ASSOCIATES WITH BETA-2.  
 CC - SUBCELLULAR LOCATION: TYPE I membrane protein.  
 CC - TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
 CC GRANULOCYTES.  
 CC - DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC - SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC - SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC - SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC - DATABASE: NAME-PROW; NOTE-CD guide CD11c entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11c.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M81695; AAA59180.1; -  
 CC EMBL; Y00093; CAA68283.1; -  
 CC EMBL; M29165; -; NOT ANNOTATED CDS.  
 CC EMBL; M29487; AAA51620.1; ALT-SEQ.  
 CC EMBL; M29482; AAA51620.1; JOINED.  
 CC EMBL; M29483; AAA51620.1; JOINED.  
 CC EMBL; M29484; AAA51620.1; JOINED.  
 CC EMBL; M29485; AAA51620.1; JOINED.  
 CC EMBL; M29486; AAA51620.1; JOINED.  
 CC PIR; A35584; RWHUIC.  
 CC HSSP; P11215; IABX.  
 CC GeneW; HGNC:6152; ITGAX.  
 CC MIM; 151510; -  
 CC InterPro: IPR000413; Integrin\_alpha.  
 CC InterPro: IPR002035; VWF\_A.  
 CC Pfam; PF00092; vwa; 1.  
 CC Pfam; PF00357; Integrin\_A; 1.  
 CC Pfam; PF01839; FG-GAP; 5.  
 CC PRINTS; PRO1185; INTEGRINA.  
 CC PRINTS; PRO0453; VWFADOMAIN.  
 CC SMART; SM00191; Int\_alpha; 4.  
 CC SMART; SM00327; vwa; 1.  
 CC PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 CC PROSITE; PS50234; VWFA; 1.  
 CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 CC Signal; Magnesium; Calcium; Repeat.  
 CC FT SIGNAL 1 19  
 CC FT CHAIN 20 1163 INTEGRIN ALPHA-X.  
 CC FT DOMAIN 20 1107 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 1108 1128 POTENTIAL.  
 CC FT DOMAIN 1129 1163 CYTOPLASMIC (POTENTIAL).  
 CC FT REPEAT 34 87 FG-GAP 1.  
 CC FT REPEAT 3 2 FG-GAP 2.  
 CC FT DOMAIN 165 351 VWFA.  
 CC FT REPEAT 402 453 FG-GAP 3.  
 CC FT REPEAT 455 517 FG-GAP 4.  
 CC FT REPEAT 518 576 FG-GAP 5.  
 CC FT REPEAT 581 633 FG-GAP 6.  
 CC FT REPEAT 633 697 FG-GAP 7.  
 CC FT CA\_BIND 466 474 POTENTIAL.  
 CC FT CA\_BIND 530 538 POTENTIAL.  
 CC FT CA\_BIND 593 601 POTENTIAL.  
 CC FT SITE 1131 1135 GEFK MOTIF.  
 CC FT DISULFID 69 76 BY SIMILARITY.  
 CC FT DISULFID 108 126 BY SIMILARITY.  
 CC FT DISULFID 655 712 BY SIMILARITY.  
 CC FT DISULFID 771 777 BY SIMILARITY.  
 CC FT DISULFID 848 863 BY SIMILARITY.

FT DISULFID 998 1022 BY SIMILARITY.  
 FT DISULFID 1027 1032 BY SIMILARITY.  
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 382 392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 697 697 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 939 939 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 480 490 G -> A (IN REF. 2).  
 FT CONFLICT 756 756 L -> D (IN REF. 2).  
 SQ SEQUENCE 1163 AA; 127885 MW; 6C4E19CC3FE2A473 CRC64;  
 Query Match 17.2%; Score 1051; DB 1; Length 1163;  
 Best Local Similarity 28.5%; Pred. No. 1.5e-67;  
 Matches 356; Conservative 201; Mismatches 474; Indels 220; Gaps 50;  
 QY 13 LVFLTGLTCLSP--FNLDHHPRLFPQPPAEFGYSVLQHVGGQRMVLGAPWDSPGDR 70  
 DB 8 LLLFTALATSLGPNLDTBELTA--RVDSAGFGDSVVOYANS--WVVVGAPQKITAANQT 63  
 QY 71 GDVYRCFVGGANAPCAKGLCDYQVGNSSHP-AVNMHLSLLETDDGCFMACAPLWS 129  
 DB 64 GGLYOC---GYSTGACE-----PIGLQVPEAVNMSTGLSTSTSPSOLLACGPVH 113  
 QY 130 RACGSVSFSGICARVADSPQGSIAPTARCPFY-MDVIVYVDGNSIYV--WSEYOT 186  
 DB 114 HECGRNMVLTGLCFILGPT-QLTORLPVSRQECPEODIYFLIDSGSISRNPATMMN 172  
 QY 187 FLRLVGLFTIDPEQIVGLVOYGESPVHEMSLGFRTKEEVARAKNLSRREGRETKTA 246  
 DB 173 FVRAVISQ-FQRP-STQSLMQFSNKQTHFFEEFRFTSNPLSLASVHQIQG-FTYTA 229  
 QY 247 QAIMVACTEGSQSGRPEEARLLVYTTDQESH-DGEELPALAKACAGVTRGIAVL 305  
 DB 230 TAIQNVVHRLPHASGARDATKILVITDKKEDSDLYXDVIYPMMAAGIRAIQVG 289  
 QY 306 GHYLRQRPDSFBEIRITIASDPDERFFVNTDEALTDIVDALGDRIFGLESHANE 365  
 DB 290 LAFQNR-----NSKELNDIASKBQEHIFVEYEDBKIDQONQKEIFALEGETTSS 344  
 QY 366 SSFGLMEOIGFSTHRLKDGLFCGVAGYDWGCVLMLEGGHRLPPRMALDEFFPALQ 425  
 DB 345 SSFELMEOEGFSAVFTDGPVLAVGSEFTV-----SGAFILYPPNMS-----PFI 391  
 QY 426 NHA-----AYLGYSVSMILKGRRLFLSGAPRFHRRKGVIAF-QLKDGAVRVAQS 476  
 DB 392 NMSQENVMDRDSYLGSTELALMKGVOSLVL-GARYOHTGKAVLFTQVSRQW--RMKAE 448  
 QY 477 LQGEQIGSYFQSELCPPLTDRDGTDLVLAAPMFLGQNKETGRVYV----- 524  
 DB 449 VTGTQIGSYFQASLCSVVDTDGSTDVLIGAPHYV--EQTRGGOVASCPLRGRRMWC 506  
 QY 525 --YLVGQOSLLTLOGLQPEPPODARFGFAMGALPDLNQGFADVAVGAPLEDGHQALY 582  
 DB 507 DAVLVGEGG-----HPW-----GRFGALTVLGDVNGKLTVDVIGAPDEEERGAIV 554  
 QY 583 LYHGTSQ-GVRPHAPQRIAAASMPHALSYFGRSVDGRLLDGDGDLVAVAGQAAILLS 641  
 DB 555 LFHGLGPSISPSHQRIAGSOLSRLOYPQOASGGGDLTQDGLVDLAVAGRGVLLLR 614  
 QY 642 SRPIVHLPSLEVPQALISVQORDRRR-GQAVCLTALCFQVTSRTP---GRWDHOFY 697  
 DB 615 TRPVLMVGSWQFIPIAETIPSAFEBREBOVSEQTLVQSNICLYIDKRKNLIGSDLOSS 674  
 QY 698 MKFTASLDEWTAGAARAFDSGQRLSPRRRLSVGNVTCEQLHPEVLDTS---DYLRPVA 754  
 DB 675 VTLIDLALDPRGLSPATQETKRN-SLSRVAVLGLKAKCE--NEVLLLPSCVDESVTBIT 731  
 QY 755 LTVTFALDNTTKP-----GPVINEGSPSTIQKLVFPSKCGGPNNECVTLVLOVNDI 807  
 DB 732 LRNLFTL--VQKPLAFENLRPMLAALQRYFTASLPEKKNGADHICQDNIGISFSFP- 788

```

OY 808 RGRKAPFVVGRRKRVVSTLE-----NRKEMAYNTSLSTIFS----- 847
DB 789 -----GLKSLVGSNLELNAEVMWVNDGDSYGT--TTFPSHAGLSRYVAE 834
OY 848 -----RNLHLASLTPORESPIKVECAAPSAHARLCSGHPVFCGTGAVTTLLEFEFSC 900
DB 835 GOKGQGLRSLHLTC-----DSAPV---GSGGTWSTSCRINHLIFRGGAQITFLATFVDVSP 886
OY 901 SSLLSQVFGKLTASSDPSLERNGLQENTNO-----TSAIOVEPHLTS----- 944
DB 887 KAVLGRL-LLTANVSSENNTPRTSKTQLELVKAVYAVVSSHEQTKYLNSESE 945
OY 945 --SESTLHRYEHPYVPG--TLVPGGPEFTTLTVQNLGCVVSGLLISALL--PAVANG 998
DB 946 KESHVAMHRYVNNIGQRLPV-----SINFWVVELNGEAVMVDVESHQ 992
OY 999 NYFLSLSQVITNNASCTIVONLTPPGPPVHPELOHTNF--LNGSNTQCVVRCGLGOL 1055
DB 993 NPSLRCS-----SEKIAPPA-SDFLAHIOKNPVLDCSIACCLFRCDVPSF 1037
OY 1056 AKGTEVSVGLRLVHNEFFRAKFKSLTVVSTFELGTEGSEVLOLTEAGRMGESLLEVO 1115
DB 1038 SVQEBIDFLKGNLSGWRQLQKXVSVVAETTFDSVYSQLPFGCAFRA-----Q 1092
OY 1116 TRPIISLMI-----LIGSVLGLLLALLLVFCIMKLGFF--AHKIPPE 1158
DB 1093 TTVLEKYVNHPTPLVIGSSIGULLLALLITAVLYKGVFPRQYKEMME 1143

RESULT 13
ITAM MOUSE STANDARD; PRT; 1153 AA.
ID AC P05555;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (leukocyte receptor MOI).
GN ITGAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88312584; PubMed=3044779;
RA Pyteia R.;
RT "Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the integrin family and an additional domain related to von Willebrand factor.";
RT Willebrand factor";
RN EMBO J. 7:1371-1378(1988).
RN [2]
RP SEQUENCE OF 11-45 FROM N.A.
RX STRAIN=BALE/c; TISSUE=Spleen;
RX MEDLINE=86287312; PubMed=2942940;
RA Sastre L., Roman J.M., Teplow D.B., Dreyer W.J., Gee C.E.,
RA Larson R.S., Roberts T.M., Springer T.A.;
RT "A partial genomic DNA clone for the alpha subunit of the mouse complement receptor type 3 and cellular adhesion molecule Mac-1.";
RT Proc. Natl. Acad. Sci. U.S.A. 83:5644-5648(1986).
RN [3]
RP SEQUENCE OF 17-28.
RX MEDLINE=85188276; PubMed=3887182;
RA Springer T.A., Teplow D.B., Dreyer W.J.;
RT "Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion glycoproteins and unexpected relation to leukocyte interferon.";
RT Nature 314:540-542(1985).
CC -1- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES. IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D

```

```

CC PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR
CC FIRINGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES
CC OF FIRINGEN GAMMA CHAIN. ALPHA-M/BETA-2 PLAY A CRITICAL ROLE IN
CC MAST CELL DEVELOPMENT AND IN IMMUNE COMPLEX-MEDIATED
CC GLOMERULONEPHRITIS. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-M
CC SUBUNIT GENE DEMONSTRATE INCREASE IN NEUTROPHIL ACCUMULATION, IN
CC RESPONSE TO A IMPAIRED DEGRANULATION AND PHAGOCYTOSIS, EVENTS THAT
CC APPARENTLY ACCELERATE APOPTOSIS IN NEUTROPHILS. THESE MICE DEVELOP
CC OBESITY.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN. ALPHA-M
CC ASSOCIATES WITH BETA-2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND
CC GRANULOCYTES.
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
CC EMBL, X07640; CA30479.1; -.
CC EMBL, M14293; AA39484.1; -.
CC PIR, S00551; S00551.
CC HSSP, P11215; 1A8X.
CC MGD; MGI:96607; Itgam.
CC InterPro: IPR000413; Integrin_alpha.
CC InterPro: IPR002035; VWF_A.
CC Pfam; PF00092; vwa; 1.
CC Pfam; PF00357; Integrin_A; 1.
CC Pfam; PF01839; FG-GAP; 5.
CC PRINTS; PRO1185; INTEGRINA.
CC PRINTS; PRO0453; VWFADOMAIN.
CC SMART; SM00191; Int_alpha; 4.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
CC PROSITE; PS50234; VWFA; 1.
CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
CC Signal; Calcium; Repeat.
CC SIGNAL 1
CC CHAIN 17 1153
CC DOMAIN 17 1105
CC TRANSMEM 1106 1129
CC DOMAIN 1130 1153
CC REPEAT 31 84
CC REPEAT ? ?
CC DOMAIN 164 350
CC REPEAT 337 400
CC REPEAT 401 452
CC REPEAT 454 515
CC REPEAT 517 575
CC REPEAT 580 632
CC CA_BIND 465 473
CC CA_BIND 529 537
CC CA_BIND 592 600
CC SITE 1132 1136
CC DISULFID 66 73
CC DISULFID 105 123
CC DISULFID 654 711
CC DISULFID 770 776
CC DISULFID 999 1023
CC DISULFID 1028 1033
CC CARBOHYD 58 58
CC CARBOHYD 86 86
CC CARBOHYD 391 391
CC CARBOHYD 696 696
CC INTEGRIN ALPHA-M.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC FG-GAP 1.
CC FG-GAP 2.
CC VWFA.
CC FG-GAP 3.
CC FG-GAP 4.
CC FG-GAP 5.
CC FG-GAP 6.
CC FG-GAP 7.
CC POTENTIAL.
CC POTENTIAL.
CC POTENTIAL.
CC GEFGR MOTIF.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC N-LINKED (GLCNAc. . .) (POTENTIAL).
CC N-LINKED (GLCNAc. . .) (POTENTIAL).
CC N-LINKED (GLCNAc. . .) (POTENTIAL).
CC N-LINKED (GLCNAc. . .) (POTENTIAL).

```

FT	CARBOHYD	772	772	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	772	772	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	801	801	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	881	881	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	907	907	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	941	941	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	980	980	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	994	994	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1022	1022	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1045	1045	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1051	1051	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1076	1076	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SO	SEQUENCE	1153 AA;	127480 MW;	178DB98BAECB0343	CRC64;

Query Match	16.7%	Score 1022.5	DB 1	Length 1153
Best Local Similarity	28.6%	Pred. No. 1.7e-65		
Matches 353	Conservative 218	Mismatches 481	Indels 181	Gaps 50

QY 13 LVFLTGLCSFENLDEHHPRLEFGRPEAEFGYSLVHVGSGQRRMLVGAPMIDGSCDGRGD 72

Dbb 7 LVLTALALCHGFNLDEHPMTFOENAKG-FGQNVV-LGGTS--VVVAAPDEAKAVNQTGA 62

```

QY      73  VYRCEVGGAHNA1PCAKGHLGD2YQLGNS3SH4-----AYNMHLMSILE5LEDDGGGFMAC6 124
      63  LYQC-----DYST7-SRCHPI8PLQV9PP10EAVNMS11LG12SLAV13STV14POOL15IL16 105
Db

```

QY 125 APLMSRACGSSVFSSGICARVDAS-FQPGSLAPRACRCPTF-MPVIVLVDGSGNSI--XP 180  
106 GPITYHONCKENTYVNGLCYFGSNNLRPPQOFPEALRECCQQOESDVIPLFDGSGSINNID 165  
Db

```
QY      181 MSEVFLERLRLVKGFLIDPEQIQ-----VGIVQYESPSPIHEWSLSDGFRTRKEEVNRAAKNL   235
          : : : : | : : : | : : : |
Db      166 FQKAKEFVSIVVM-----EQFKSKTLFSIMQYDSEFRHNFPTNDFKRNPSPRHVSPRI   218
```

0y 23 SRREGRETTAQAIWVACTGFSOHGGRBEARALLVWTVDGESHDGEELPAALKACEAG 295  
Db 219 KQLNGR-TTTSAGIRKRVRELFHKIKNGARENAKILVITDGEK-GDPLDYKVIPEAD 276

Qy 296 R--VTRYGAVLGHRLRQDPSSFLREIRIASDPDERFEFNVIDEALTDIDVALGR 353  
277 RAGIRIVLGV-GNAFNK---PQS--RRELDITASKPAGEHVFQVNDFEALNTIQNLOQEK 331  
Db

0Y 354 IFGLGSHAHENESSPGLEMOQIFGTHRLKDGIETGHWGVAYDM -GGSVLWLEGGHRLFP 412  
 332 IFALEGTQRTGTSSTSEHHEMSEGGFASITNSGPNPLGVSQFDMAGCAFLYTSKDAYTFIN 391  
 Db

QY 413 RMALEDEFPALONAAAYLCYSVSSMILLRGRRLTSGARFRHNGKVIAPQLKKDGAVR 472  
          |      |      |      |      |      |      |      |      |      |  
Db 392 TTRVDSDM-----NDAYLGYA-SAVILNRKVQSIVLGARFRYQHIGLVNMR-ENFGTWIE 443

473 VASQSLQEQIGSYFSESLCPDITDSDGTTVLVVAPEMFQPNKKEGRVYVYLQG--- 529  
 444 PHTSIKGSQIGSYFGASLSCVMDMDADGNTNMLIGAPHY--EKRRGGQGVSVCPPLRGA 501

530 --QSLLTLCSTLQEPPODPAFGEFAMGALPDLNQGFAVDVAVGAPLDEHGKALVLYHGT 587  
502 RMQGEALLHGDD-QGHR--MGRFGAALTVLDEVDNCKLTDLVAIGAPGEGENQGVAVIFYGA 558

Db 559 SIASLSASHSHRIICAHFSPGLQYRGQSLSGSKDLTMDGLMDLAVAGAGQHLLLRAPVL 618

```

OY 647 HLTPLSELEVPQALISVQQRDQRR-----GQEAVALTALCFQVTSRTGGR--WDHQF 696
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 619 RLKATMEFSPKKVARSVFACQEQVLKNKDAGEVRACTL-----RRKNTDRLREGDIQS 672

```

QY YMRRTASLDIEWTAGRAAFEGSGGRLSPRLRLSGVNTCEQLHFVLD-TSDYRFPVAL 755

673 TVTVDLALDPVRSRIRAFDET-KNTRTRRTQVFGLMGKCKETLKLILPDCVDSDVSPIL 731

OY        TATFALDNTKTP-----GVLNKGSPSTIQIKLVPFSSKCGPDNECVTDLTQLVNMDIR     808  
         : : : : :  
756        LRVLRSLRQPLPGLMADQRFFTFAMPEFKNGNDSDICOD--LSTITMSAM     787

Dd        RLNVTLT---VGEPRLRSFGNLRLPVLMADQRFFTFAMPEFKNGNDSDICOD--LSTITMSAM     787

```

OY      809 GSRKAPFVWGGRKKVULVSTTLERKKNAYNTSLIIFFSNHL-----ASTTPRESEPIK 864
Db      788 G-----LDTLVVGGPDGFENMSVTLNRKDGSDSYGTQVTVVYPSGLSTRKSDASONPLTKKWPWF 844
OY      865 VECAPAPSA-----HARLCS-----VGHPIFOGAKATFLLFEF--SCSLLSQVFGK-LTAS 914
Db      845 VKPARESSSSSEBGHALKSTMTNINHPIFPANSEVTFNVTVDVDSHAFCKNLLKALVAS 904
OY      915 SDSLERNGTLOENTQTSAYIOYEPHLLFSS-ESTL-----HREYVHPY 957
Db      905 ENMSRFT--HKTFQLELPVKVKAIVIVTSDSESSIRYLNFTASEMYSKVIOHQYOFNNL 961
OY      958 G--TLPPGPBPBEFTTLRQNLGCVVYVSGLLISALLRAVAGNGYFLSLSQVITNNASCI 1015
Db      962 GQRLAPV--GVEMPIVQJNNVTYWDHPQVIFSQNLSASCH-----1000
OY      1016 VQNTPEPPGPVPH---EELQHTNRNLGNSVTOCCVAVCHGLQGLAKGTEVSGLRLVHNE 1072
Db      1001 -----TEQKSPF-HSNFRDQLERTFVLNCSVAVCKRIQCDLPSEFNTQEIPIFNTLKNLSFD 1055
OY      1073 FFRRAKRSKLTJVVSFELGTEEGSVLQLTASRNSLSLLEV-VQTRPILISLWLTGSVL 1131
Db      1056 WYIKTSHGHLILVASTEILFNDSAFALLPQOESYVRKSTETKVEPVENHPVPLVIGSSI 1115
OY      1132 GGLLLALLVFCMLKLGFPFAHKKIPEEKREBEK 1164
Db      1116 GGLVLLALLTAGLYKLGFF-----KKQYK 1139

```

RESULT 14	
ITAE_MOUSE	
ID ITAE_MOUSE	STANDARD;
	PRT; 1167 AA

DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Integrin alpha-E precursor (Integrin alpha M290)

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus  
OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.  
RC STRAIN=AKR;  
RX MEDLINE=95187992; PubMed=7882170;  
RA Smith T.J., Ducharme L.A., Shaw S.K., Parker C.M., Brenner M.B.,

RT "MORINE M290 INTEGRIN EXPRESSION MODULATED BY MAST CELL ACTIVATION.",  
 RL IMMUNITY 1:393-403(1994).  
 CC -I- FUNCTION; INTEGRIN ALPHA-E/BETA-7 IS A RECEPTOR FOR E-CADHERIN. IT  
 CC MEDIATES ADHESION OF INTRA-EPITHELIAL T-LYMPHOCYTES TO EPITHELIAL

CC SUBUNIT GENE EXHIBIT A MARKED REDUCTION IN THE NUMBERS OF  
CC INTRAEPITHELIAL LYMPHOCYTES IN THE GUT AND IN THE DEVELOPMENT OF  
CC GUT-ASSOCIATED LYMPHOID AGGREGATES, SUPPORTING A SPECIFIC ROLE FOR  
CC THIS INTEGRIN IN MEDIATING RETENTION OF LYMPHOCYTES IN THE

CC -1- SUBUNIT, HETEROIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAINS LINKED BY / DISULFIDE BOND. ALPHA-E ASSOCIATES WITH BETA-7.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.  
CC -1- SIMILARITY: CONTAINS 7 EG-CAP REPEATS  
CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.  
CC WITH 1-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSER1) IS A VFMA DOMAIN.  
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSER1) IS A VFMA DOMAIN.  
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSER1) IS A VFMA DOMAIN.

CC This SMIS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way





```

Db 984 -----HGENLFGAVFOL-----QICVPIKLOPQIVRVNLTQD----- 1019
Oy 1028 HPEELQHTRLNSNTQCCVVRCH-----LGLANGTEKSV-----LIR-----L 1068
Db 1020 HTECTOSQAPACSDVQVHKEMHSVCAITSNKENVTAAELSVHTQQLANDSELPI 1079
Oy 1069 VANEPRFAKFKSL-----TVSTPELGTGEGSVQLTEASRWSSELLVOTREPLIS 1122
Db 1080 LGEISNKLVEBLSNENRKTITVFLKEE-----TR-----S 1114
Oy 1123 LWLIGSVLGLLLALVFCMLKGFPAHK--KIDPEEKREKLE 1166
Db 1115 LPLIGSSIGGLLVVIAIILFKCGFFPKYQQLMESTRAQLK 1160

RESULT 15
ITAE HUMAN STANDARD; PRT; 1179 AA.
AC P38570; Q9NZU9;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-E precursor (Mucosal lymphocyte-1 antigen) (HML-1
DE antigen) (CD103 antigen) (Integrin alpha-IEL).
GN ITGAE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
ON (1)
RX SEQUENCE FROM N.A., AND SEQUENCE OF 19-38 AND 179-188.
RC TISSUE=Lymphocytes, and Leukemia;
RX MEDLINE=94164962; PubMed=8119947;
RA Shaw S.K., Ceppek K.L., Murphy E.A., Russell G.J., Brenner M.B.,
RA Parker C.M.;
RT "Molecular cloning of the human mucosal lymphocyte integrin alpha E
RT subunit. Unusual structure and restricted RNA distribution.";
RL J. Biol. Chem. 269:6016-6025(1994).
RN [2]
RP REVISIONS TO 88-114.
RA Parker C.M.;
RN Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 53-1179 FROM N.A.
RC TISSUE=Fetal kidney;
RX MEDLINE=20138496; PubMed=10673275;
RA Touchman U.W., Ankster Y., Dietrich N.L., McDowell G.,
RA Shoclerstuk V., Bouffard G.G., Beckstrom-Sternberg S.M., Gahl W.A.,
RA Green E.D.;
RT "The genomic region encompassing the nephropathic cystinosis gene
RT (CTNS): complete sequencing of a 200-kb segment and discovery of a
RT novel gene within the common cystinosis-causing deletion.";
RN Genome Res. 10:1165-1173(2000).
RN [4]
RP MUTAGENESIS OF ASP-109 AND PHE-316.
RX MEDLINE=20400502; PubMed=10837471;
RA Higgins J.M.H., Cernadas M., Tan K., Irie A., Wang J.-H., Takeda Y.,
RA Brenner M.B.;
RT "The role of alpha and beta chains in ligand recognition by beta 7
RT integrins.";
RL J. Biol. Chem. 275:25652-25664(2000).
RN [5]
RP FUNCTION: INTEGRIN ALPHA-E/BETA-7 IS A RECEPTOR FOR E-CADHERIN. IT
RN MEDIATES ADHESION OF INTRA-EPITHELIAL T-LYMPHOCYTES TO EPITHELIAL
RN CELL MONOLAYERS.
CC [6]
CC SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAINS LINKED BY A
CC DISULFIDE BOND. ALPHA-E ASSOCIATES WITH BETA-7.
CC [7]
CC SUBCELLULAR LOCATION: Type I membrane protein.
CC [8]
CC TISSUE SPECIFICITY: EXPRESSED ON A SUBCLASS OF T-LYMPHOCYTES KNOWN
CC AS INTRA-EPITHELIAL LYMPHOCYTES WHICH ARE LOCATED BETWEEN MUCOSAL
CC EPITHELIAL CELLS.
CC [9]
CC DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WFPA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.

```

```

CC [10]
CC SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC [11]
CC SIMILARITY: CONTAINS 1 WFPA DOMAIN.
CC [12]
CC SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC [13]
CC DATABASE: NAME=PROW; NOTE=CD guide CD103 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd103.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL: L25851; AAB59359.2; -.
DR EMBL: AF168787; AAF43107.1; -.
DR HSSP: P1215; IABX.
DR Genew: HGNC:6147; ITGAE.
DR MIM: 604682; -.
DR InterPro: IPR000413; Integrin_alpha.
DR InterPro: IPR002035; VWF_A.
DR Pfam: PF00092; vwa; 1.
DR Pfam: PF00357; Integrin_A; 1.
DR Pfam: PF01839; FG-GAP; 4.
DR PRINTS: PR01185; INTEGRINA.
DR PRINTS: PR00453; VWFADOMAIN.
DR SMART: SM00191; Int_alpha; 4.
DR SMART: SM00327; VWA_1.
DR PROSITE: PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE: PS50234; WFPA; 1.
KW Integrin; Cell adhesion; Receptor; Polymorphism; Magnesium; Calcium.
KW Signal; Repeat.
FT SIGNAL 1 18
FT CHAIN 19 1179 INTEGRIN ALPHA-E.
FT CHAIN 19 177 INTEGRIN ALPHA-E LIGHT CHAIN.
FT DOMAIN 19 1124 INTEGRIN ALPHA-E HEAVY CHAIN.
FT TRANSMEM 1125 1147 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1148 1179 POTENTIAL.
FT DOMAIN 181 198 CYTOPLASMIC (POTENTIAL).
FT REPEAT ? ? GLU-RICH (ACIDIC).
FT REPEAT ? ? FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT REPEAT ? ? FG-GAP 3.
FT DOMAIN 145 199 X-DOMAIN (EXTRA DOMAIN).
FT DOMAIN 200 391 WFPA.
FT REPEAT 401 456 FG-GAP 3.
FT REPEAT 457 506 FG-GAP 4.
FT REPEAT 510 571 FG-GAP 5.
FT REPEAT 573 638 FG-GAP 6.
FT REPEAT 641 693 FG-GAP 7.
FT CA_BIND 522 530 POTENTIAL.
FT CA_BIND 586 594 POTENTIAL.
FT CA_BIND 654 662 POTENTIAL.
FT SITE 1150 1154 GPRFR MOTIF.
FT DISULFID 70 79 BY SIMILARITY.
FT DISULFID 126 159 BY SIMILARITY.
FT DISULFID 706 762 BY SIMILARITY.
FT DISULFID 823 829 BY SIMILARITY.
FT DISULFID 893 907 BY SIMILARITY.
FT DISULFID 1008 1033 BY SIMILARITY.
FT DISULFID 1041 1057 BY SIMILARITY.
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 211 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 934 934 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 954 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1065 1065 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1096 1096 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 360 360 D -> E.
FT /FTID=VAR_008884.

```



```

FT  VARIANT  1041  1041  C -> S.
FT  MUTAGEN  208  208  /FTID=VAR 008885.
FT  MUTAGEN  316  316  D->A: LOSS OF E-CADHERIN BINDING.
FT  CONFLICT  477  477  F->A: LOSS OF E-CADHERIN BINDING.
FT  CONFLICT  482  482  V -> I (IN REF. 3).
FT  CONFLICT  950  950  O -> R (IN REF. 3).
FT  CONFLICT  950  950  R -> W (IN REF. 3).
FT  CONFLICT  1019  1019  A -> V (IN REF. 3).
SQ  SEQUENCE  1179 AA; 130088 MW; E558902BDF995E1 CRC64;

Query Match      15.3%; Score 932.5; DB 1; Length 1179;
Best Local Similarity 26.8%; Pred. No. 5.4e-59;
Matches 348; Conservative 188; Mismatches 451; Indels 303; Gaps 58;

QY  8 HTLFLPLVLTGLCSFNLDEHHRPLP--GPPEAEFGYSVLQH--VGGQGRMWLVGAPWD 63
DB  5 HTLLCLASL-ALLAANVDVARPWLTPKGA P---FVLSLSLHQDSTNTQMTLWLTSP--- 58
QY  64 GPGSGRRGVYRC-----PVGAMNAPCAKGLHDYQLGNSHPAVNMHGLMSLL 113
DB  59 -RTKRTGPGPLHRCSTLVODEILCHPV---BHVPIPKGR-----HRGVTVV 98
QY  114 ETDGDCGFMACAPLMSRAGS--SVFSSGICARVDAFQPGS-----LAPTA---- 159
DB  99 RS--HHGVLTICQVLVRPHSLSELTGCSLGPDLRPOAQANFPDLLENLDPPARVDT 156
QY  160 -----ORCPYMDVIVLIDGNSIYP- 180
DB  157 GDCYSKEGGEDDVTARQRALKEKEBEDEKEBEDEBEAGETIAIILIDGSSIDPP 216
QY  181 -MSEVOTFLRLVGLKFIIDPEQIOVGLVOYGESPVHEWMLGDPRTKEEVVRAKILSRRE 239
DB  217 DFORADPFISSNMNRNFEKCFECNFAIVQYGGVITQTEFPLRDSQDMASLARVQNT-QY 275
QY  240 GRETKAQAIVNACTEGFSQSHGGRPEARLLVYVTDGE--SHDGEELPAALKACEAGRYT 298
DB  276 GSVTKTASAMOVULDSITSSHGRKAKKAMVVLTDGIFEDPLMTTVINSPPKQGV 335
QY  299 RYGIATLGHYLRORDPSSFLREIRTIASDPDEFNFVNTDEAALTDIALDGRIFGLE 358
DB  336 RFAIGV-GEFFKSART---ARELNLIASDPDETHAFKTYNMALDGLSKLRYNISME 390
QY  359 GSHAENESSFGLEMSOIGFSTRKLD-GILFGMGAIVDMGSSVLMLEGHRLPPPRMA-- 415
DB  391 GTVGD---ALHYQLAQIGFSAQIILDERQVILGAVGAFDWSGAL-----LYDTRSRRG 440
QY  416 --LEDEFPALQNHAA---YLGYSVSSMLLRGRRRLFLSGAPRRFRHKVIAFOLKKG- 469
DB  441 RFLNOTAAADAADAQAQSYLYGAAV-AVLAHTCSSLSYVAGAPQYKHHGAV--FELQKEGR 497
QY  470 AVRVAQSLQGEQIGSYFGSELCPDLTDRTDGTVDVLLVAAPMFLGPQNKETGRVYVYLVQ 529
DB  498 EASFLPLVLEGRQMGSTFGSELCPVDIDMGSTDFLLVAAPFY--HYHGEGRVYVYRLE 555
QY  530 Q-SLLTLOGLQPEPP-QDARFGFAMGALPDINODGFADVAVGAPLE-----DGHQ-QAL 581
DB  556 QDGSFSLARILSGHPCFTNARFGFAMAAGDLSODKLTDVAIGAPLEGGADGASFGSV 615
QY  582 YLYHGTQSGVRPHPRQRIAAASMPHALLSYFGRSVDGRLLDDGDDLVADVAVGAQAAILLS 641
DB  616 YLYNGHMDGLSASPSORIRASTAVAPGLQYFGMSMAGGFDISGDGLADITVGTGQAVFR 675
QY  642 SRPIVLTPLSLVTPQALISVQDRCRRQEAVALCTLAALCFQVTSRTPGRM--HOFVVR 699
DB  676 SRPVVRLKVSMAFTSALPI-----GFNGV-VNVRLCFEISSVTTSBSGLAREALLN 726
QY  700 FTAASLD-----EYTAARAAFDGSGQRLSPRRLLSVGNVTCQ 738
DB  727 FTLDVDVVGKORRRRLQCSDVRSCLGCLREWS-----SGSOLCEDLLIMPTGELCE 777
QY  739 LHFHVLDTSDYLVRVALTVTFALD---NTTKGPVLINE-GSPTSIQKLVPSKDCGPRN 793
DB  778 -----DCFSNASVKSIVQLQTPREGQTDHQPILDHRYTEFPALFQI-PYEKACKKNKL 827

```

```

QY  794 ECVTDLVLQVMDIRSKRAFPVVGRRKVLVSTLLENRKENAYNTSLIIFSRNLHA 853
DB  828 FCVAELQIATTVSQGE-----LVVGLTKELTLNINLTNSGSDSYMTSMALNYPNLTQK 881
QY  854 SLTPRBSPIKVECAAPSAHARL---CSVGHDPV-QGAKVTFLLERFSGCSLLSOVF 908
DB  882 RM--OKPPSPNQCDPPQPVASVLIMNCRIGHPLVKRSSAHVSVMQLE-----ENAF 932
QY  909 GKLTSQSLERNGTLOENTAQTSAYIOY-----EPHLF--SSESTLHRYE--VH 955
DB  933 PNRITDITVYTNSNERSLANETHTLQFRHGFVAVLSPSIVNVNTOGSLHKEFLPH 992
QY  956 PVGTLVPVGPGEFFKTLTRVONLGCYVSGLIISALLPAVAGNYFLSLQVIT----- 1009
DB  993 VHGKULF--GAEYQIQICVPT---KLKGLVAAVKKLTRQASIVCTWQSGERACAYSSV 1046
QY  1010 -----NNACTIV---ONLTPEPGPVHPE-ELQHTNRLNGSNTQCCVVRCHLQOLAKG 1058
DB  1047 QHVEEMHSVSCVIASDKENVT-----VAAEISMDHSEELLKDYVTELOI---LGEISFN 1096
QY  1059 TEVSQGLRIYHNE-----FPRAKFKSLTVVSTFELGTEBGSVLQLEASRWSSLELV 1113
DB  1097 KSLYEGLAENHRTKITVFLDEKYNHL----- 1125
QY  1114 VOTRPIILSLWILIGSVLGLLLALLVFCIMKLGFFAHL 1153
DB  1126 -----PIIK-----GSV-GGLLVILVILVIFKCGFFARK 1155.

```

Search completed: July 16, 2003, 07:52:48  
 Job time : 36.365 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:52 / Search time 59.9093 Seconds  
(without alignments)  
1872.646 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106

Sequence: 1 MELPFTVTLPLVFLTGLC.....GFFAHKKIPREEKREKLEQ 1167

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	1894.5	31.0	1180	2 A35854	integrin alpha-1 c
2	1887	30.9	1151	2 A45226	integrin alpha-1 c
3	1756	28.8	1181	2 A33998	integrin alpha-2 c
4	1754	28.7	1170	2 I45914	integrin alpha 2 s
5	1744	28.6	1178	2 S44142	VLA-2 protein homo
6	1105	18.1	1170	2 S03308	cell surface glyco
7	1097	18.0	1163	2 I56126	lymphocyte function
8	1085.5	17.8	1153	1 RWHUB	cell surface glyco
9	1054	17.3	1163	1 RWHUB	cell surface glyco
10	1022.5	16.7	1153	2 S00551	leukocyte surface
11	932.5	15.3	1179	2 A53213	integrin alpha-E c
12	738	12.1	1035	2 I58409	integrin alpha-9 c
13	732.5	12.0	1054	2 J07294	alphan integrin -
14	690	11.3	1039	2 A41131	lymphocyte-feyers
15	684	11.2	1038	2 S06046	integrin alpha-4 c
16	661	10.8	1041	2 T31437	integrin alpha-1 c
17	655	10.7	1041	2 A55348	integrin alpha-1 c
18	640.5	10.5	1137	2 J05950	integrin alpha-3 c
19	602.5	9.9	1051	2 A40021	integrin alpha-6 c
20	600	9.8	1091	2 A41543	integrin alpha-6 c
21	598.5	9.8	1072	2 B36429	integrin alpha-6 c
22	596.5	9.8	1073	2 A38457	integrin alpha-6 c
23	594.5	9.7	1135	2 T16186	alpha-7 integrin -
24	584.5	9.6	1044	2 T10050	integrin alpha-v c
25	581.5	9.5	1053	2 I55514	VLA-3 alpha subuni
26	576.5	9.4	1044	2 S16516	integrin alpha-8 c
27	576.5	9.4	1051	2 A35761	cell surface glyco
28	574	9.4	1048	2 A27421	integrin alpha-5 c
29	573.5	9.4	1106	2 S38783	integrin alpha cha

30	570	9.3	1045	2 S60571	integrin alpha v c
31	566	9.3	1053	2 S44250	integrin alpha-5 c
32	563.5	9.2	1034	2 A36108	integrin alpha-V c
33	562	9.2	1049	2 A27079	fibronectin recept
34	543	8.9	1146	2 S40311	integrin - fruit f
35	540	8.8	1039	2 A34269	integrin alpha-2b
36	495	8.1	1037	2 A60163	glycoprotein IIb -
37	485	7.9	1226	2 S44824	glycoprotein IIb -
38	475	7.8	1394	2 A29677	position-specific
39	445.5	7.3	1139	2 S28277	hypothetical prote
40	438	7.2	1086	2 T18523	integrin alpha cha
41	425.5	7.0	1115	2 T09403	integrin alpha cha
42	425.5	7.0	1115	2 T09433	integrin alpha cha
43	394.5	6.5	191	2 I47230	VLA-2 protein - pi
44	374	6.1	764	2 I36916	glycoprotein IIb -
45	303.5	5.0	604	2 I36917	glycoprotein IIb -

ALIGNMENTS

RESULT 1					
A35854					
Integrin alpha-1 chain precursor - rat					
C:Species: Rattus norvegicus (Norway rat)					
C:Date: 23-Oct-1990 #sequence_revision 13-Sep-1991 #text_change 20-Sep-1999					
C:Accession: A35854; S11243					
R:Rignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto, J.					
U:Cell Biol. 111, 709-720, 1990					
A:Title: Molecular cloning of the rat integrin alpha-1 subunit: a receptor for laminin a					
A:Reference number: A35854; PMID:90338125; PMID:2380249					
A:Accession: A35854					
A:Status: preliminary					
A:Molecule type: mRNA					
A:Residues: 1-1180 <ICN>					
A:Cross-references: GB:X52140; NID:G56493; PID:CAA36394.1; PID:G56494					
A:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology					
A:Keywords: cell adhesion; cytoskeleton; transmembrane protein					
F:170-345/Domain: von Willebrand factor type A repeat homology <vWA2>					
Query Match					
Best Local Similarity 31.0%; Score 1894.5; DB 2; Length 1180;					
Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;					
QY	13	LVPLTGLCPNLTBHHRLPPPEAEFGYSVIOHVGGRWMLVGA PMDPSGDRRGD	72		
DB	19	LTVLILGFCVSNVIVKMSMFSGVEDMFGITVQVNEBESKWLIGSPLVGQPKARTGD	78		
QY	73	VYRCPEVGAHNA PCAKGLHDYQLGNSHPAV-----NMHLGMSLLETGDCGFWACAPL	127		
DB	79	VYKCPVGERAMPCKVDLP-----VNTSIPNVTEIKENMTRG-STIVTNPNGFLACPL	133		
QY	128	WSRACGSSVSSGICARVDASFQPGSLAPLAQCPTMYDVIVLDGNSITYPMSSEVOTF	187		
DB	134	YAVYCGHLHYTTGICSDVSPTFQVNSFPAP-VQECSTQLDIVILDGNSITYPMSVIAF	192		
QY	188	LRLIVGKFLIDPEQIOVGLVYGESPVHEMSLGPRTKEEVYRAKNLSREGEETKQAQ	247		
DB	193	LNDLKRMDIDPKQIQVIGIVYGENVTHEFNLKSYSTEEVLVAANKIGRGGGLQTMIAL	252		
QY	248	AIMVACTEGFQSGHGRPEEARLLVVTGESHDEELPAALKACEAGRVYRGIAYLVGH	307		
DB	253	GIDPARKEAFTFARGARGAVKKNVITVDGSHNNYRKQVIOGCEBDNIORFSAITLGH	312		
QY	308	YLRQRDPSFLREIRTIASDPDERFPFNVTDALTDIVDALDRIFLGEGSHAEENS	367		
DB	313	YNRGNLSTEEKVEERIKSIASEPTEKHFFNVSDDELATVIVALKERIFALPATADQSAAS	372		
QY	368	FGLMSQIGFSTHLKGLLFGWVGAYDMGSGVLMLEGGHRLFPFRNALDEFPALQNH	427		
DB	373	FEMENSQTFSAHYSQDMVMIGAVGAYDMNSTVMQKANKMNVIPHNTPFQTE--PAKNE	430		
QY	428	--AAVLGYSVSSMLRGGRRLFLSGAPRFRRHGVIAFQLKKGDAVVAQSLQGEQIGSY	485		

```

Db      431 PLASLYGYVNSATIPGD-VLYIAGQPRVNHGTQVVIYKKA-EDGNINILQTLGGEQIGSY 488
Qy      486 FGSELCPLDTRDRCCTDVLIIAAMPFLGPONKETGVVYVYLVGQOGLTLTQGLQF----- 541
Db      489 FGSVLTITDIDKOSYTDLLLVGAPMYWGEKEQGVYVAV-NQTRFEVQMSLEPIROT 547
Qy      542 -----BPQDARFGFAMGALPDINOGFADVAAGVLEDDHOGALYLYHG 586
Db      548 CCSGLNDNSCTKENKNEPCGARFGTAAVAKDLNDVGFNDVIGALEDDHAGAVIYHG 607
Qy      587 TOSGVRPHAPQRIAAASMPHALSYFGRSVDGRJLDGDDILVDVAAGAAILLSRPV 646
Db      608 SGKTIEAAYQRIIPSGDGGKTLKFPGOSIHGEMDLNGDGLTDTYIGLGAALFMARDVA 667
Qy      647 HLPLSEVTPQAISVVOODCRRGQAEVCLTALCQVTSRTPGRMDHGFYMFASLDE 706
Db      668 VAKVTNFEENKVNICKKCRVEGKETVCINATMCHVKKSKEDSIYEADLOYRVTLDS 727
Qy      707 WTGAAAPADSGQRLSPRLRLSVGNVTCQQLHFVLDTSDYLRPAVLTVPALDNTTK 766
Db      728 LRQISHSFSGTQERIKQR-NITVASECIRHSFYMDKHDFQDSVRTLDF--NLTD 782
Qy      767 P--GPVLNBSPTSIKLVFESKDCGPBNECVTDVLQVNMDIRSKKAPFVVRGGRRV 824
Db      783 PENGVPVLDALPNSVHEHIFPAKDCGNKERCISDLTLNVST--TEKSILLVKSQHDRF 838
Qy      825 LVSTLENKENVNYSLSIIFSRNLHLASLTP-QRESPIKVECAPSAHARLCSGHV 883
Db      839 NVSLYKNNKGSVYNTVYVQSHPNLIFSGIEIIOQDS-----CESNONITTCRGYPP 891
Qy      884 FQTGAQVTEFLLEFEFSCSSLSQVFGKLTASSDLSLRNGTLQENTAQTSAYIOYEPHLLF 943
Db      892 LRAGEIVTEFKIIQFNTSHSENAIIHLSTSDSEEPLESINDENEVNIIPVKEVGLQF 951
Qy      944 SSESILHRYEHPDYGLP-----VGPPEFKTTLFVQNLGCVVSGLIISALLP-AVAV 996
Db      952 YSASASHHSIVANETIPEFINSTEDIGNEINVFYIRKRGHPMDELSTISFPMULTAD 1011
Qy      997 GGNVFLSLSQVITNNASCIVONLTBPFG-----PVPHEPELOHTNRALNGSNQCVVAC 1050
Db      1012 GYVPLVPIGMSSSDVNCPRSLDEDFGINSKKMTISSEVAKGTTIODCSTCGVATI 1071
Qy      1051 HLGLQAKG-TEVSVGLRLVHNEFFRRARFKSLTVSTFELGTEESVQLTEASHWSBS 1109
Db      1072 TCSLPSDSLQVNVSL--LLMKPTFIRAHFSSNLTLRGEIKSENS-LLTSSNNKRL 1128
Qy      1110 LLEAVOT-RPLISLWILIGSVLGLLLALLVFCMLKLGFAHKKIPEEKREE 1163
Db      1129 AIOISKDGLGRHVPLVWIIISAFAGLLMLLTLALMKIGFF--KRPLKKMKEX 1180

RESULT 2
A45226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #ext_change 31-Dec-2000
C:Accession: A45226
R:Blotsewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Experimental source: hepatoblastoma cell line HepG2
A>Note: sequence extracted from NCBI backbone (NCBI:124326)
C:Superfamily: unassigned
C:Domain: von Willebrand factor type A repeat homology <WAL>
F:142-317/Domain: von Willebrand factor type A repeat homology <WAL>

```

Query Match 30.9%; Score 1887; DB 2; Length 1151;  
 Best Local Similarity 36.3%; Pred. No. 8.8e-133;

```

Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;
Qy      23 FNLDEHRLPFGPEAREGYSVLQHVGGGRWMLVGA-PWMDPSGDRGRDVRCPVGAH 82
Db      1 FNVADKSNSTSGVPEDMFGITVQOYENEBGKMWLIGLPLVGPKNRTGDIYKCVGGE 60
Qy      83 NAPCAKGLGDYOLGNSSHPAV-----NMHLGMSLLEFDGDFMACAPALWSRACSSVF 137
Db      61 SLPCVKTDLR-----VNTSIPNTEVKENMTFG-STLVTNPNPGFLACCPVLYRGGHLYH 115
Qy      138 SSGTCARVADASTPQGSILAPTAQRCPTVMDVYIVLDGNSNIYPMSEVOTFLRLVGLFI 197
Db      116 TTGISDSVPFQVNVNSIAP-VQECSTQDLDTIVLDGNSNIYPMQSVAFNLDLKNDI 174
Qy      198 DEQIOLGVQVQGESEPHWMSLGDRTKEEVYRAKNLSREGREXTQAQAIMVACTGEF 257
Db      175 GPKQTOVGIVQGEVNTHEPNLKNSSFEELVVAANKIVQGRGRTMTALGDTARKAF 234
Qy      258 SOSHGRRPEAARLLVVVTVDGESHDEELPAALKACEAGRVTRYCIAVLGHYLRQRDPSS 317
Db      235 TEARGARGVKKVWVITVDGESHDRRLKVIYQDCEDENIQRFSSIALIGSVNRGLSTEK 294
Qy      318 FLREIRTIASDDEFFFNVTDEALTDIVDALGRITGLGSHANENSSFGLEMSQGF 377
Db      295 FVEEIKSISABETEKGFFNVSDDELVTITVITLGERIPALBATAQOASAASEMEWSQGF 354
Qy      378 STHRLKDGILFCMVAGYDMGSLWMLBEGHRLFPFRMLLEDFPPLQNHAAVLGYSVS 437
Db      355 SAHYSQDMVMGAVGAYVMNGTVWQKASQIIPNTTFNVESTKNEPLASYLGITYNS 414
Qy      438 MLRGARLFLSGARFRHGRKVIAFOLKQDGAHVAAQLOGEOIGSYFGEELCPDTR 497
Db      415 ATASSGDVLYIAGQRYVNHGTQVVIYRM-EDGNIKIQLTSGEQIGSYFSGILTTDIDK 473
Qy      498 DQTTVLLVAAMPFLGPONKETGRVYVYLVGQOGLTLTQGLQF----- 541
Db      474 DSNTDILLVGAAPMYVGTKEQGVYVAV-NQTRFEVQMSLEPIKOTCCSSROHNSCTT 532
Qy      542 -----BPQDARFGFAMGALPDINOGFADVAAGVLEDDHOGALYLYHGTCGVRPHAPQR 598
Db      533 EKNPEPCGARFGTAAVAKDLNDVGFNDVIGALEDDHAGAVIYHSSGKTIRKEYAQOR 592
Qy      599 IAAASMPHALSYFGRSVDGRJLDGDDILVDVAAGAQAAILLSRPVHLTPSLSEVTPQA 658
Db      593 IPSGDDGKTLKFPGOSIHGEMDLNGDGLTDTYIGLGAALFWSHRDVAVVKTMFBNK 652
Qy      659 ISVVRDQRRRQGEAVCLTALCQVTSRTPGRMDHGFYMFASLDEMTAGARAFQGS 718
Db      653 VNIQKNHMEGKETVCINATVCEVKKLSKEDTIEADLOYRVTLDSLRQISRSFSGT 712
Qy      719 GORLSPRLRLSVGNVTCQQLHFVLDTSDYLRPAVLTVPALDNTTKP--GPVLNBSGP 776
Db      713 QERKQQR-NITVASECTKHSFYMDKHDFQDSVRTLDF--NLTPENGVPVLDSDLP 767
Qy      777 TSIQTLVPSKDCGPBNECVTDVLQVNMDIRSKKAPFVVRGGRRKVLVSTLENREN 836
Db      768 NSVHEYIIPAKDCGKKEKISDLSLHV---ATTEKDLIIVRSQDKFNVSILVNTKDS 823
Qy      837 AYNLSLSIIFSRNLHLASLTPORESPIVVECAAPSAHARLSVGHFPVQTAQVFLLEF 896
Db      824 AYNNTTIVHSPNVLVFSI-----EAIQDSC--ESNHNITTKVGPFLRRGEMVFKILF 877
Qy      897 EFGCSLSQVFGKLTASSDLSLRNGTLQENTAQTSAYIOYEPHLLFSSSESTLHRYEHP 956
Db      878 QPNTSYLMEVNTIYISATSDSEEPETISDVNVANISIVKKEVGLQFSSASEVHISTAA 937
Qy      957 YGTLP-----VGPPEFKTTLFVQNLGCVVSGLIISALLPAVAGN---YPLSLSQV 1007
Db      938 NETVEVINSNTEIDGNEINIFYLIRKSSGFPMPPELKISIPNMTNSNGVPLVPTGLSS- 996
Qy      1008 ITNNASCIVONLTBPFG-----PVPHEPELOHTNRALNGSNQCVVACHGLQAKKTEV 1061
Db      997 -SENANCRPHIFEDPFSINSKKMTTSDHLKRGITLDCNTCKFATINCNTLS-SDISQV 1054

```

```

Oy 1062 SGLLRVHNNEFRRAKRSKLTIVSTFPLGTEBGSVLQLTASRMSSELELVOT-APIL 1120
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1055 NVSL--ILMKPFIKSYSSNLTIRGELREKMS-LVLSSSNQKRELAIQISKDLPGR 1111
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1121 ISLWILIGSVLGGLLLLALVFCMLKLGFFAHKKIPEEKKEE 1163
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1112 VPLWVILLSAFAGLLMLLMLLIALMKIGF---KRPLKKKKEK 1151
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 3
A33998
integrin alpha-2 chain precursor - human
N:Alternate names: CD89b; platelet glycoprotein GPIb; VLA-2/collagen receptor alpha-2 ct
C:Species: Homo sapiens (man)
C:Date: 30-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 21-Jul-2000
C:Accession: A33998; B56793; A53117
R:Takada, Y.; Hemler, M.E.
J: Cell Biol. 109, 397-407, 1989
A:Title: The primary structure of the VLA-2/collagen receptor alpha(2)-subunit (platelet
A:Reference number: A33998; MUID:89308679; PMID:2545729
A:Accession: A33998
A:Molecule type: mRNA
A:Residues: 1-1181 <TRAX>
A:Cross-references: GB:J17033; NID:G33906; PIDN:CAA4894.1; PID:G33907
A:Note: the authors translated the codon GAT for residue 802 as Gln, GTC for residue 800
R:Catmell, B.; Parmantier, S.; Leung, L.L.; McGregor, J.L.
Biochem. J. 279, 419-425, 1991
A:Title: Separation of important new platelet glycoproteins (GPIa, GPIc, GPIc*, GPIIa and
A:Reference number: A56793; MUID:92061944; PMID:1953640
A:Accession: B56793
A:Molecule type: protein
A:Residues: 30-43 <CAT>
A:Experimental source: platelet
R:Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Tsung, Y.L.; Gafford, A.
J. Biol. Chem. 269, 463-469, 1994
A:Title: The human alpha-2 integrin gene promoter. Identification of positive and negative
A:Reference number: A53117; MUID:94103255; PMID:8276836
A:Accession: A53117
A:Molecule type: DNA
A:Residues: 1-16, 'V', 18-21 <ZUT>
A:Cross-references: GB:L24121; NID:G400342; PIDN:AAA1619.2; PID:G4583535
A:Note: authors translated the codon GFA for residue 17 as Leu
C:Genetics:
A:Gene: GDB:ITGA2; CD49B
A:Cross-references: GDB:128031; OMIM:192974
A:Map position: Sq11.1-Sq11.2
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
C:Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-1133/Domain: extracellular #status predicted <EXT>
F:112-347/Domain: von Willebrand factor type A repeat homology <VWA>
F:1134-1154/Domain: transmembrane #status predicted <TM>
F:1155-1181/Domain: intracellular #status predicted <CYT>
F:105,112,343,432,460,475,699,1057,1074,1081/Binding site: carbohydrate (Asn) (covalent)

Query Match 28.8%; Score 1756; DB 2; Length 1181;
Best Local Similarity 34.3%; Pred. No. 6.2e-123;
Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

Oy 11 LPLVFLTL-----CSPFLIDENHRLFGCPREAEKRGVYLQHVGGGQRMVLGAPWD 63
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 11 LPLLLVLLASGILNCCLAIVNGVLEAKIFSGPSESGVAVQGFNPKGMWLVGSPMS 70
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 64 GPSGDRCDVYRCPYGGAHNAFCAGKHLDGYOLGNSSHPAVNHLGMSLLETDDGGFM 122
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 71 GFPEHRMDVYKCPV-DISTATCEKLNLOTSTSPNVTEMKTNNLSGLILTRNMGTGFL 129
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 123 ACAPLMSAGCSVSSSGICARVUDASPDQGSILPTARCTYVDVYVLVDGSSYIPWS 162
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 130 TCGPLMAQQCGQYTTTCGSDISPDQLSFSFATPPCSSLIDVVVVVCDENSSITPMD 189
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 183 EVQGFRLRVGKFLIDPEQIQVGLVQGESPVHNSLDPFTKEVVRPAKNLSRRGREG 242
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db	190	AVKMFLEKVOGLDIDFTPTQVGLQYANPRVVFNNLTYYTKKEMIVALSQTSQYGGD	249
Qy	243	TKTAQAIMVACTEGFSQSHGCRPEABRLVWTDGESHGELPALKACEAGVTRYGI	3020
Db	250	TNTFGAIQYARKVAYSAASGRRSATKVMVVVTDGESHGSMLEAVIDQCHNDILAFGI	3090
Qy	303	AVUGHYLRORDDSSFLKIRITIASDPDBRFFPNVTDEALTDVLDGRIPELES	362
Db	310	AVUGYINRNALDTPKNIKEIKAIASIPERYEPFNVSDEALLERAGLTGEOIFIS	368
Qy	363	ENSSFGLEMSQIGFT--HRKDLILGCMGAYOMGWSYLM--LEGHRLEPPRMAL	419
Db	369	QGGDNTOEMESQVGFADYSQNDILMLGAVCAFEMSGTIVOKTSHGLLFP----	423
Qy	420	FPBALQ--NHAALGYSVSMRLGGRRLFLUSGAFRPHRKGVALFOLKKGAVRVA	477
Db	424	FDQILDRHNSSTVIGSVVA--ISTESHFVAAGAPRAYTOQIYLVSNEGNITVIA	482
Qy	478	QGEQIGSYFSGELCPDLTDGDTTVLLVAAPMLGPNKETGRVYVY----LV	532
Db	483	RGDQISYFSGVCSVDVDKDTITTVLLVGAAPMWSDLKKEGRVYVYFTTIK	542
Qy	533	LTIQGLPBPPODAFPGFAMALDLMQDGFADVAVAPRIEDHGQALYIYHGT	592
Db	543	--DEG--PEGIENTFGAIALALDMMGDENDVIVGSPLENSGAVIYNHOGTIR	597
Qy	593	PHAPORIAAA--SMFALSYFGRSDGRDLDDGDLVVAAGAQAAILSSRPVLT	650
Db	598	TKYSQKILSGDGAFRHLOYFRSLDGYDGLNGDSITVSIAGAGQVQYOLMSO	651
Qy	651	SLEVTPOAISVYORDCRRRGOEAVCLTALCFQVTSRTPGRMDHQFYMRFTAS	710
Db	658	EASFTPEKITLVKNAQ-----IILKLGSAKFR--PTKONQVAIVANITL	704
Qy	711	ARAPFGSGORLSPRLL-----RLSVGNV-----TEQULHFVILDTSYL	758
Db	705	-----DGFSSRYVSRGLFKENNERCLOKRMVNVNAQSCPEHIYIOEPS	759
Qy	759	FALDNTTKG--PVLNBSPTSIOQLVPEFSKDCBPNCVLDIYQNMDRGSKAP	816
Db	760	ISLEN--PETSALANSETAKVFSIPPHKDCGEGDLCSIDLVDVR-QIPAAQ	815
Qy	817	VREGRRKVLVSTLENRKENAVNTSLIIFSRNHLASLTFQRESPI--KVEC	872
Db	816	VSNQNRRLPFSVTLKKKRESAYNTGIVDFSENLFPSFS---LPVDGTEVTC	871
Qy	873	HARLCVGHVPOTGAKVTFLLSEFSCSSLSIOYFGKLTASDLSLERNGTLO	922
Db	872	KSAVCQVGPALRREQOVFTINFPDNLONQNASLSFQALSSEOEENKA--	929
Qy	933	AYIOYERPHLLFSESTLRYEVRPGTLR-----VGRPEFKTTLAVONIG	966
Db	930	IPLLYDAEHLRSTNINFEYISSDGNVPSIVHSFEDVGPFISLKV--TTG	988
Qy	987	ISALLPVAHAGNVFPLSQVITNNASCIVONLTPP-----GPVHPBELO	1030
Db	989	VIIHIFQYTKENRPLMYLTGVOTDRAGDISCADINPLKIGQTSYSSVFS	104
Qy	1039	NGSNTQCOVVRCHLQOLAKGTSEVSGLLRLVHNEFFRAKFSLTJVSTEL	109
Db	1049	NCRTASCNVTCWLKDVHMKGEYFVNVTTRIMNGFPASTQYQVQLAAAE	110
Qy	1099	QLTEASRMSESLEVVQTRPILISMT-----LIGSVLGGLLALLVYC	114
Db	1109	-----VIEDNTVTIPIIMPKPDEKAEVPTGVLIGSIAGILLALLVA	115
Qy	1146	KLGFPAHK-----KIPBE 1158	
Db	1155	KLGFPAKRYKRYKPNDE 1172	

145914  
 integrin alpha 2 subunit - bovine (fragment)  
 C.Species: Bos primigenius taurus (cattle)  
 C.Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 20-Sep-1999  
 C.Accession: I45914  
 R.Kamata, T.; Puzon, W.; Takada, Y.  
 J. Biol. Chem. 269, 9659-9663, 1994  
 A.Title: Identification of putative ligand binding sites within the I-domain of integrin  
 A.Reference number: A54402; PMID:94193647; PMID:7511592  
 A.Accession: I45914  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: mRNA  
 A.Residues: 1-1170 <KAM>  
 A.Cross-references: GB:I25886; NID:9439695; PIDN:AAB5925.1; PID:9439696  
 C.Superfamily: unassigned collagens; von Willebrand factor type A repeat homology  
 F161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 28.7%; Score 1754; DB 2; Length 1170;

Best Local Similarity 33.9%; Pred. No. 8.6e-123;

Matches 411; Conservative 235; Mismatches 465; Indels 102; Gaps 30;

```

9 LFLPLVFLTLGL---GSPNLDEHRLFPPEAEFGYSVLQHVGGGQRMVLVGAHPDGP 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2 LQVLVFSOGILNLCVAHVNGLPKAKIFSGPSSEQGYAVQVQFINPKNMILLVSGPWSGF 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66 SGDRGDVYRCVPVGAHANAPCAKGLH-GDYOLGNSHPAVNMHLSLETDDGGGFMAC 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 PKRMMDVYKCPV-DLSTTCEKLNQTSMSNVTDEMTNMSLGLTLRNVTGSGFLTC 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 APLWRACGSSVFSSGICARVDASFPQGSAPTAQRCPTMDVIVLIDGNSIYPMSEV 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 GPLMAQCGSQYVTTGVCSDPDQLRTSFAPAVQTCPSFIDVVVVCDSNSIYPMDAV 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 QTFRLRVKGLFIDPEQIOVGLVQGESPVHEWSLDFRTKEVVAANKLSRBERETK 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
181 KNFLKFFVQGLDGPRTKTMGLIQYANRPVFNLTFFSKBEMIKATSQTFYGGDLTN 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 TQOAIWVACTEGFSQSHGRPEARLLVVTVDGSHDEGLPALKACAGATRYGIAV 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 TPKALVYADATVASTAAGRPKATKMWVTTDGHSDGSLKAVIDQCKNDLIRGIAV 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305 LGHYLRORQDPFSLREIRTIASDPERFFFNVTDEAALTDVALDGRIFGLESHAEN 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 LGVFNNAALDTKNLKEIKAIASIPTEHFVNVSDEADLLEKAGTIGEQFISIEGT-VQG 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 EESFGLNMGIOGSTHRLDGL-ILFGMCAVWGSVLM-LEGHRLPPRALEDERP 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 GDNFQWEMSQVGSABYSPPQNNILMLGAVGAYDMSGTIVQKTPHGLIFS-----KQAF 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422 PALQ--NHAAYLGYSVSMILRGGRFLSGAPRFHRRGVIAFOKKCGAVVAQSLQG 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
415 QILQDPNHSYLCYVAS--ISTGNSVHFVAGAPRAYTQIVYVSNENGVTVIOSQSG 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480 EGIQSYFSGELCPDLTRDQDTDLVLAAMPFLGPONKETGRVYVVLVGOQSLLTLQGLT 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
474 DQIGSYFGSVLCAVDVNDKDTITDVLLVGAPMVYNDLKEGGRVYLFITIKGILNMHGFLE 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
540 QEPPEPDATFFPAMGLPDLNDGFPADVAVGAPLEDGHGALYLYHGTOSGVRPHAGTI 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
534 GNGLENARFGSIALISIMNDGFNDVIQVSPLENQSAVYIYNGHCGMRLRYSOXI 593
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 AASAMPHA--LSYFGSVYDGRDLDDDDLVAVVAGOGAAILLSRPVYHLFSLSEVTPQ 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
594 LGSDFRFSHLYQYFGSLDGYDLNDSITDVSVGFQGVVQVLMGSIADVSADAFTRK 653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 AIVSVQRDRRRRQGEAVCLTALCFQVTSRTPEGRMDHGFYMRFTASLDEWTAG---ADA 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
654 KITLTKMAE-----IKTKLCFSAKFR-PTNQNNUVAIVYNTIDDOFSSRVISRG 704
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
714 APDGSQRLSPRLRTSVGNVTCQGLHFVLDTSVLRPALVTALDNTTKRG--PVL 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 LFKENNERCLQKTMIVSOAO-RCSEYI IHIQEPSDIISPLNLCMNISLEN--PGTNPAL 760
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 772 NEGSPTSIQKLVFPFSKDCGPDNECVTDLVQVNMIDIRSGRAKAPFVVRGRKVLVSTLE 831
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 761 EAVSTVAVFSLFPFKDGDGDCVCSIDLIVNV-QQLPRTQGPFTVSNQNKRLTFSVQLK 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 832 NKENAVNTSLISITSRNLHLASLTPORESP---KVEG-AAPSAHALCSVGHVFOTG 887
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 820 NKKEBAYNTEIVDSENLFFPASFWS---MPDGTETVCOJLASQKSVTCNVGYPALMSK 875
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 888 AVVTLLPEPFSKCSLSQVGVKLTAFSSDSLEKNTGLQENVAQTSAYVQYEPHLLFSSSES 947
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 876 QCVTTTFNDFELWLNQASISFPALSESQEN--MADNSNKLSTLYDEHITST 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 948 TLHRYE-----VHPYGLTPVGPPEFKTLTVQNLGCVVSGHLLSALPAVAG 997
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 934 NINPFVSLDGVSSVHVSFEDI---GPKFISIKV-TTGSVPVSMASVLIHIQYTKD 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 998 GNYFLSLSQVITNNA---SCIVQNLTEPPGPPVHP-----EELQHNRLNG 1040
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 989 KNPLMYLTVGHTDQAGDISCEAE-----INPLKIGOTSSSVFSKSENRHAIKELNC 1039
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1041 SNTQGVVRCHLGOLAKGTEVSVGLRLVHNEFPRAKFKSLTVVSTFELGTEEGSVLQ 1100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1040 RTASCNSIMCWRDLQVKGVEFLVNSTRIWNGTFASTFQYVQLTAABEDITYNQIYI 1099
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1101 TEASRWSLSLEVQTRP---ILISLWILIGSVLGLLALLVFCMLKGFPAHK--- 1153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1100 EE---NVTYIPLTIKMKHEKEVPTGVIVSGVINGILLALLVIMKLGFFKRYEM 1155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1154 -KIPEEKREKL 1165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1156 ANPDEPDETTEL 1168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

# RESULT 5

S44142

VLA-2 protein homolog - mouse

C.Species: Mus musculus (house mouse)

C.Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Sep-1999

C.Accession: S44142

R.Jedelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Danjanovich, L.; Latze

submitted to the EMBL Data Library, January 1994

A.Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not

A.Reference number: S44142

A.Accession: S44142

A.Status: preliminary

A.Molecule type: mRNA

A.Residues: 1-1178 <EDB>

C.Cross-references: EMBL:229987; NID:9473098; PIDN:CAAB2877.1; PID:9473099

C.Superfamily: unassigned collagens; von Willebrand factor type A repeat homology

F169-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 28.6%; Score 1744; DB 2; Length 1178;

Best Local Similarity 35.3%; Pred. No. 4.9e-122;

Matches 420; Conservative 221; Mismatches 479; Indels 70; Gaps 31;

```

9 LFLPLVFLTLGL---GSPNLDEHRLFPPEAEFGYSVLQHVGGGQRMVLVGAHPDGP 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10 LILQILMVQGLNLNLAINVGLPKAKITSGSSSEQGYAVQVQFINPKNMILLVSGPWSGF 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 66 SGDRGDVYRCVPVGAHANAPCAKGLH-GDYOLGNSHPAVNMHLSLETDDGGGFMAC 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70 PENRMGDVYKCPV-DLPATCEKLNQTSMSNVTDEMTNMSLGLTLRNVTGSGFLTC 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 APLWRACGSSVFSSGICARVDASFPQGSAPTAQRCPTMDVIVLIDGNSIYPMSEV 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 129 GPLMWHQGNQYVATGICSDVSPDQLRTSFAPAVQTCPSFIDVVVVCDSNSIYPMDAV 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 QTFRLRVKGLFIDPEQIOVGLVQGESPVHEWSLDFRTKEVVAANKLSRBERETK 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 189 KNFLKFFVQGLDGPRTKTMGLIQYANRPVFNLTFFSKBEMIKATSQTFYGGDLTN 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 245 TQOAIWVACTEGFSQSHGRPEARLLVVTVDGSHDEGLPALKACAGATRYGIAV 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      249  TFRALFADVAVYSGTSGRPGATKVVVVVTDSHDSKLTQVIOQCNDEILRFGLAV 308
Qy      305  LGHYLRQRPDPSSFLREIRTTIASDPERFFFNVTDEALTDIVDALGRIFGLESHAEN 364
Db      309  LGYLRNALDITNLIKELIKALISTPERKFFRNVADEALILEAKGLIGREIFIEGT-VQG 367
Qy      365  ESSFGLKMSQIGPSTHRL--KDGILFGVAVDWMGCVLMLEGGHR-LFPPRMALDEDF 420
Db      366  GDNFGEMMAQVGFSDVAFQNDILMLGAVGAFDMGSLIV-QETSHKPVIFP-----KQAF 421
Qy      421  PRALQ--NHAALVGVSVSSMLRGGRRLFLSGAPFRHGRKXIAIOLKDKGAVRVAQSLQ 478
Db      422  DQVLDRNHSFLGSAVAISTEDGVH-FVAGAPPANTYGOVLVSVKQGVNTVYIOGR 480
Qy      479  GEOIGSFGESELCPLDTRDGTTPDVLVAAPMFLPQNKETGRVYVYVVGQOSLTLTLOT 538
Db      481  GQOIGSFPSVLCSDVDVDKDTITDVLVGAFTYMDLKKBEKVLVFTTKILMLQHQFL 540
Qy      539  LQPEPPDARFGFANGALPDMLNODGFADYVAGAPLEDHOGALVLYHGTQSGVPRHPAQR 598
Db      541  EGPFGTGNARFGSALIALSDINMDGFNDYVGSFVENENSGAVYIYNGHQITRTKYSQK 600
Qy      559  IAAA--SNPHALSYGRSDVGRLLDDGDVLVAVNAGAACALILSSRPVHLTPELEVP 656
Db      601  ILGSGAFRRRLQFFGRSLDYGDLNGDSITVSGALGOVQLWMSOSIADVAIALFLTP 660
Qy      657  QAIIVQDRRCRRGQEAVALCTALCFQVTSRTPGMDHQFYVRFASIDEMTAGARAADF 716
Db      661  DKITILNDK-----ITKLCFRAEFRRAGQ--NNQVALLFMTITDADSHSRVSR 711
Qy      717  GSGQRLSPRRLR--LSVGNV-TCEQLHFFVLDTSYLRVALVTFTALDNTTKPG--PVL 771
Db      712  GVFRNSERFLQKMMVNVNEVKCSSEHHSIQKPSDVVNPDLDRVIDISEN--PGTSPAL 768
Qy      772  NEGSTSTOKLVNPFKDCGPNECYTDVLQYNNMIRSKRAPPVYVGRKRVAVSTLE 831
Db      769  EAVSETVAVKFSIPFYKECGSDGICISDILVD-QOLPALQTSFIVSNQNKRLTFPSVLTK 827
Qy      832  NRKENAVNTSLIIFSRNLHLASLTPORESPI--KVECAAPSARAL-CSVGHVFPQNG 887
Db      828  NRGESAINTVLAERSENLFPSFS---MPVDGIEVTECVSSQKSVTCDVGYFALISE 883
Qy      888  AKVTELFEEFSCSILSGVPEKLTASPSLSERNGLQENTQTSAYIOYEPHLLFSSES 947
Db      884  QCVETFINDFVLQNLQQAALNFQAFSESQETNNA--DNSVSLTPIPLYADEMLTST 941
Qy      948  TLHREVHPYGLP-----VGPPEPKTLFVQNLGCVVSGLIISALLPAVHAGVYF 1001
Db      942  NINFEISSDENAPSVIKSVEDIGPKFIFSLKV-TAGSAPVSMALVTTHIPQYTERKQPL 1000
Qy      1002  LSLSGVITNNA---SCI--VQNLTEP---PGPVHPELOHNTNRLNGSNTQGVVRCHLG 1053
Db      1001  LYLTEGIQDQAGDICTAEINPLKLPHTAPSVSPKXENFRHKEIDCRTTSCSNITCWK 1060
Qy      1054  QLAKGTEVSGLLRLVHNEFFRRARFKSLTVVSTELGTEGSGVQLQTEASRMSLELV 1113
Db      1061  DLHMAKEVFINVTVMVNRTPAFSTFQYQLTAAAEIDITHNQ-LFVIEENAVTIPLMIM 1119
Qy      1114  VQTRILISLMLIGSVLGGLLALLVFCMLKLGFF-AHKKI--PBE 1158
Db      1120  KTEKAEPVPTGISIIAGILLLLAMTAGMKLGFFKQYKQKMGQNPDE 1169

```

## RESULT 6

Cell surface glycoprotein CD11a precursor - human  
 N:Alternate names: leukocyte adhesion glycoprotein LFA-1 alpha chain; leukocyte function C:Species: Homo sapiens (man)  
 C:Date: 28-Feb-1990 #sequence revision 28-Feb-1990 #next\_change 20-Aug-1999  
 C:Accession: S03308; A47458; A48765; A48765; S36044  
 R:Larson, R.S.; Corbi, A.L.; Berman, L.; Springer, T.  
 J:Cell Biol. 108, 703-712, 1989  
 A:Title: Primary structure of the leukocyte function-associated molecule-1 alpha subunit

```

A:Reference number: S03308; MUID:89139587; PMID:2537322
A:Accession: S03308
A:Molecule type: mRNA
A:Residues: 1-1170 <LAR>
A:Cross-references: EMBL:Y00796; NID:931421; PIDN:CAA6747.1; PID:931422
A:Note: part of this sequence was confirmed by protein sequencing
R:Corneill, R.D.; Gollahan, K.A.; Hickestein, D.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 4221-4225, 1993
A:Title: Description of the leukocyte function-associated antigen 1 (LFA-1 or CD11a) pro
A:Reference number: A47458; MUID:93248261; PMID:8097887
A:Accession: A47458
A:Molecule type: DNA
A:Residues: 1-20 <CNR>
A:Note: sequence extracted from NCBI backbone (NCBI:130862, NCBI:130863)
R:Shelley, C.S.; Farokhzad, O.C.; Arnaout, M.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 5364-5368, 1993
A:Title: Identification of cell-specific and developmentally regulated nuclear factors t
A:Reference number: A47565; MUID:93281759; PMID:8099450
A:Accession: A47565
A:Molecule type: DNA
A:Residues: 1-20 <SHE>
A:Cross-references: CB:M95609
R:Nueda, A.; Lopez-Cabrera, M.; Vara, A.; Corbi, A.L.
J. Biol. Chem. 268, 19305-19311, 1993
A:Title: Characterization of the CD11a (alphaL, LFA-1alpha) integrin gene promoter.
A:Reference number: A48759; MUID:93374910; PMID:8103515
A:Accession: A48759
A:Molecule type: DNA
A:Residues: 1-20 <NUE>
A:Cross-references: EMBL:Z22804; NID:9311405; PIDN:CAA80461.1; PID:9311406
A:Genetics:
A:Gene: GDB:ITGAL; CD11A
A:Cross-references: GDB:119757; OMIM:153370
A:Map position: 16p11.2-16p11.2
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C:Keywords: cell adhesion; cytoskeleton; glycoprotein; heterodimer; surface antigen; tra
F:1-25/Domain: signal sequence #status predicted <SIG>
F:154-1170/Product: leukocyte adhesion glycoprotein LFA-1 alpha chain #status predicted <
F:154-317/Domain: von Willebrand factor type A repeat homology <VMA2>

```

## Query Match

```

11 LPLVFLGL-----CSPNLDENHRLPPGPEA--EGVSVLQHYGGGQRMVLVGAPWD 63
9 MAMLLSGFFFPADASTYNDLVGRARSP--SPFRGRHGRVLD-VGNG--VIVGAGE 63
64 GPSGDRRGDYVRCVGAHNPACAKGHLGDYQLGNSHPAVNMLGMSLETDDGCFMA 123
64 GNS---TSLVQCGSGTCHCLPVT-----LQGSNV--TSKYLGMTLATDPTDGSILA 110
124 CAPLMSRACGSSVFPSSGICARVDSFQ--PGGSLAPTYQRC--PTYMDVYIVLDGNSIYP- 180
111 CDPGLSRCDQNTVLSGLCYFRQNLQPMQGRPGFOECIKGVNDVLFDFGSMISQPD 170
181 -MSEVFLRLVYKLFIDPEQIOVLVOGESVHMSLGF--RTKEVNVAAAKLS 236
171 EFQKILPMKVMKKL--SNTSYQPAVQFSTYKTEDESDYKMDPDLHLGVKML 228
237 RREGREFTQAQIIVACTEGFSQSHGRPEARLLVVVTDGSHDEGLPALAKACEAGR 296
229 LL-----TNTPAIVVATEVREELGARPDATKYLIIITDSEALDSGNIDAKD----- 278
297 VTRYGLAVLGHYLRQRPSSFLREIRTIASDPERFFFNVTDEALTDIVDALGDRIFG 356
279 IIRYIIGKRFQYKEQET-----LHKFASKPASEPKLIDTPEKLDLFTLQKXIYV 333
357 LEGSHAENESFGLEMSQIGPSTHRLKDLIFGVAVDWMGCVLMLEGGHRLLPPRMAL 416
334 IEGTSKODLTSFNNELSSGISADLSRGHAYVGVAKDWAAGFLDKADLIQ----- 385
417 EDER-----PRALQNHAAVLYGVSVSMILRGGRRLFLSGAPFRHGRKXIAIOLKDKGA-V 471

```

```

Db      386 DDTFIGNELTPREVRAGVLYTWTMLPSRKQKTSLSASGAPRYQHMGRVLLFQEPQGGHM 445
      472 RVAQSLQGEIQISYFSSSELCPLDTRDGTDDVLLVAAPMFLGQNKETGRVYVY--LVG 528
      446 SOVQITIHGQIGSYFSGELCGVDVDDGETELLILAPFYEQQ--RGRVFPYQRRQGG 503
      529 QOSLLTLQGTLOEPPODARFGFAMGALPDLNODGFADVAADVAGAPLEDGQGLYLHNGTO 588
      504 FEESVSLQG--DGYVL-GRFGAITALDLINDGVLVDVAVGAPLEE--QGAAYITNGNH 558
      589 SGVRPAPQRIAAASMPHALSYFGRSVDGRLDLGGDLVDVAVGAQGAAILLSRPYHL 648
      559 GGLSPQPSQRIEQTQYLSGIQWFGRSIHGVKDLLEGGLADVAADVAGASQMIVLSSRPVDM 618
      649 TBSLETPQALISVQVQDC---RRQGEAVCLTALCFQVSTSPRKMHOFRMR----- 699
      619 VTLMSFSPAELIPHEVECSYSTSNKKKEGVNIT--ICFOIKSLYP-----QFQGLVAVNL 671
      700 -FTASLDEWTAGARAFAFDGSGQRSLSPRLLSVGNVTCQLHFNH-VLDTSDYLRPVALTV 757
      672 TITLQDGHGRTRRRGFLPGGRHNL--RNNIAYTMSCTDFSHFPVCVQDLISPTNVSL 729
      758 TVAL---DNITYKPGVNLNBSPTSIGKL-----VPFSKDCGPDNECVTDLVLQVN-MD 806
      730 NFSLMEEGETPRDQRAQKQDIPILRPSLHSETWEIPEFKNCGEDKCKCANLRVSPSPAR 789
      807 IGSRRAPFVNGGRKVLVSTLLENKENAVNTSLIIFSRNLHLASLTQR-ESPITV 865
      790 SRALRLTAPA-----SLSVELSLNLEDDAYVQDLHFPGLSFRKVMELKPHSQIVP 843
      866 ECAPASAHARL-----CSGVHPVQFGAKVTFLFEFSSCSLSQVFGKLTASS--DS 917
      844 SCEELPEERLILSRALSCVSSPIFKAGISVA-----LQMMNTLVNBSWGBS 891
      918 LERNGT-----LQNTAQTSAVIOYEPHLF--SSESTLHRYEVNRYGTLFVPGC 965
      892 VEHNAVNTCNNEEDSLNEDNSATTIIPILYPINILIQDEBDSITLY-----VSFTPKGP- 944
      966 PEKKTLLRQNLGCVYVSGLIISALLPAAVAGNRYLSLSQVITNNASCIQVNLTPRP-- 1023
      945 -----KIHOVKMYOVRIPQSI-HDHN-----IPTLEA---VVGVPQPSSE 981
      1024 GP-----PVHPEELQHTNRLNGSNTOC---QVVRCHLGQLAKGTEVSGLLR 1067
      982 GPTTHQMSVQMEPRVYCHEDLE--RLQDAAPRCPLGALFRCPVFRQELILVOVIGTLE 1038
      1068 LV---HNEFPRAKFKSLTVVST--PELGTEEGSVLQLTASRWSSESLLEVQTRPILI 1121
      1039 LVGELEASMFSLCSLSLISFNSSKHFHLYGSNVASIAQVY-----MKVDVYVEKQMLY 1091
      1122 SLWILIGSVLGGILLALLLVFCMLKGFPAHKKIPEEKREKLE 1166
      1092 -LYVLISG--IGGLLELLLIPIVLYKGVFP-----KRNLIKEME 1126

```

## RESULT 7

```

156126 lymphocyte function-associated molecule-1-alpha - mouse
C:Species: Mus musculus (house mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 20-Aug-1999
C:Accession: I56126
R:Kaufmann, Y.; Tseng, E.; Springer, T.A.
J:Immunol. 147, 369-374, 1991
A:title: Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit
A:Reference number: I56126; NUID:91268576; PMID:2051027
A:Accession: I56126
A:status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1163 <RES>
A:Cross-references: GB:M60778; NID:q198785; PIDN:AAA9426.1; PID:q198786
C:Genetics:
A:Gene: LFA-1

```

C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo F:151-315/Domain: von Willebrand factor type A repeat homology <VWAl>

```

Query Match      18.0%; Score 1097; DB 2; Length 1163;
Best Local Similarity 28.4%; Pred. No. 1.7e-73;
Matches 356; Conservative 192; Mismatches 470; Indels 234; Gaps 48;

      12 PLVFTTGE-----CSFPLNDEHHNPLFPPEAEERGVVLQVHGCGGRMMLVGAPWDPSP 66
      8 PRLLILGLQIPAKMSYVLDTRPQSLAAGRHNGVQLDIEDG---VVVGAPGE--- 60
      67 GDRGDVYRCVPGGAHNAACKAGHLDYQLGNSSHPAVNMHLGSLTDDGGFMACAP 126
      61 GDNLTGLYHCTRSSSEFCQVPS-----LHGSNH--TSKYLGMTLADAAKGSLLADCP 110
      127 LMSRACSSSVSSGICANVDSFQPGSLAPTAQRCPY-----MDVVVLQSGNSI- 178
      111 GLSRTCDONTYLSGLC---YLFPQSLGFMQLQRPAYQCMKGVLDVLFVLDGSGSLD 165
      179 -YPMSEVOTFLRLVGLKFLIDPEQIQVGLVQGESPVHEWSLGPRTKEEYVRAKANLSR 237
      166 RKDFEKILFPMKDVVRKL--SNTSYQFAAVQFSTDCTREFFLDY-----VKQNKNDV 217
      238 REGRE-----TKTQAIWVACTEGFSQSHGGRPBARLLVVTVDGESHDBELPALKA 291
      218 LLGSVQPMFLTLNTRFRAINYVAVHFKESGARPDATKVLITDGEASDKGNISA--- 273
      292 CEAGNVTVGAVLGHVYRQRDPSFLREITIASDDEEFNFVNDDEALTLTDVDAIG 351
      274 -ADHTYITIGIGHFVSVQKQ---KTHIFPSEVEEFVKILDTFEKLDLFTDLO 326
      352 DRIFLESASHAENESSFLEMISOIGFSTRLLKDLGLFGMVAYDMGGSVLMEGHRLLP 411
      327 RRIYIBETNODLTSFMELSSSGISADLSKHAIVGAVAKWAGGFLDLR----- 379
      412 PRMALED-----EPPPALQNHNAVIGYVSSMLRGGRRLLSGAPRRRHGKYIAF 463
      380 -----EDLQATFVQGEELTSDVREGYLYTVAMWTSRSSRPLLAAGAPRYOHQVQLLF 434
      464 QL-KKQAVNRAVQSLQGEIQISYFSSSELCPLDTRDGTDDVLLVAAPMFLPQNKETGRV 522
      435 QAPENQGRMNQTKIEGTQISYFSGELCSVDLDQGAELLLGAPLFPEEQ--RGRV 492
      523 VYVILVQOSLLTLQGTLOPEPPD--ARFGFAMGALPDLNODGFADVAADVAGAPLEDGQGL 581
      493 FTY-QRQSLFEMWSELQDGYPLGRGAALITLTDINGRLDVAVAGAPLEE--QCAV 549
      582 YLVHGTQSGVRPAPQRIAAASMPHALSYFGRSVDGRLDLGGDLVDVAVGAQGAAILLS 641
      550 YIFNGKPGGLSPQSORIQGAQVPPGIRFWFGRSIHGVLDLGGDRLADVAADVAGEGVVLS 609
      642 SRPIHLPLSLLEVTPQALISVQVQDC---RRQGEAVCLTALCFQVSTSPRKMHOFR 697
      610 SRPVADVATELSFSEELIPHEVECSYARBEQKHGVLYKA--CFRIKPLRP-----OFQ 662
      698 MR-----FTASLDEWTAGARAFAFDGSGQRSLSPRLLSVGNV-----TCEQLHFNH-VL 744
      663 GRLLANLSYTIQDGHRRSRGFLPPDGSHEL-----GNTSITPKSCLDLFFHFRPI 714
      745 DTSGLRVALVTTPALDNTTKPGVNLNBSPT-----SIQKLVPSKCD 788
      715 CIQDILSPINSLNLSL-----LEEEGTPROQKGRAMOPILRPSIHVTKEIPFEKN 766
      789 CGPDNECVTDLVLQVNMDIRGSRKAPFVVRGGRKVLVSTLLENKENAVNTSLIISFR 848
      767 CGEDKKCANLTL-----SPARSGPLRMS-ASLAEWLTLSNGEDAPVAVRLDLDFPR 820
      849 NLHLASLTP-QRESPIKVCAAPSAHARL-----CSGVHPVQFGAKVTFLFEFSSCS 901
      821 GLSFRKVMQLPHSRMPVSCSELTEGSSLLTKLKCVSSPIFKAGQSVS----- 870
      902 SLLSQVFGKLTASS--DSLERNGLT-----QENTAQTSAVIOYEPHLFSSSESTL 949

```



Db 871 --LQMFNTLNSNDEPELNGTCHCENENSLOEDNSAHHPIVLYVNIITTEQEN- 927  
 Qy 950 HREYHVPYGTLPVGGPGEFFKTLRYONT-----GCYVSGLIISALLPAVAHGQNYPL 1002  
 Db 928 -----STLYISFPTKPGPKTQOVQVHVYVRIQPSAVDHNNPTLEALVGVPRHSEDDI 979  
 Qy 1003 SLSQVITNN--ASCIYQNTTEPPGPPVHVEELQHNRLNGSNTQCVV-RCHLGQLAKT 1059  
 Db 980 TYTWVQVQDPLVTCHESEDKRRSSSEABQP-----CLPGVQRCPIYFRMELLQVGT 1032  
 Qy 1060 EVSVGLRLVHNEPFRRAFKS-LTVVSTFEELTEGVSVLQLTASRMSES---LLEVV 1114  
 Db 1033 V-----ELSKETASSTLSICSLSPNSSKRFHL-YGSKASEAVLYKVDLI 1080  
 Qy 1115 QTRPILISLWILIGSVLGLLALLVFLVCMKLGFFAHKKIPEEKREKLE 1166  
 Db 1081 HEKEML-HYVVLGSG--IGGLVLLPLFLALYKVGFF-----KNNLKEKME 1122  
 RESULT 8  
 RHWUB  
 cell surface glycoprotein CD11b precursor [validated] - human  
 N.Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac-1  
 C.Species: Homo sapiens (man)  
 C.Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text change 31-Dec-2000  
 C.Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; 152567  
 J.Corbil, A.V.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.  
 J. Biol. Chem. 263, 12403-12411, 1988  
 A>Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b)  
 A.Reference number: A31108; MUID:88315033; PMID:2457584  
 A.Molecule type: mRNA  
 A.Residues: 1-1153 <COR>  
 A.Cross-references: GB:J03925; NID:9187284; PIDN:AAA59544.1; PID:9307148  
 A.Note: part of this sequence was confirmed by protein sequencing  
 R.Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.  
 J. Cell Biol. 106, 2153-2158, 1988  
 A>Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mac-1  
 A.Reference number: A28915; MUID:88257215; PMID:2454931  
 A.Molecule type: mRNA  
 A.Residues: 1-999, 501-965, 'P', 967-1153 <ARN>  
 A.Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; PIDN:AAA594  
 A.Note: the authors translated the codon TAC for residue 1129 as Thr  
 R.Shelley, C.S.; Arnaout, M.A.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991  
 A>Title: The promoter of the CD11b gene directs myeloid-specific and developmentally regulated expression  
 A.Reference number: A41600; MUID:92073318; PMID:1683702  
 A.Molecule type: DNA  
 A.Residues: 1-9 <SHR>  
 A.Cross-references: GB:M6724; NID:9180018; PIDN:AAA58410.1; PID:9553215  
 R.Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
 A>Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion receptor  
 A.Reference number: A41933; MUID:88190151; PMID:2833753  
 A.Molecule type: mRNA  
 A.Residues: 917-1042 <AR2>  
 A.Cross-references: GB:M18044  
 R.Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989  
 A>Title: cDNA sequence of the alpha subunit of the human neutrophil adherence receptor  
 A.Reference number: A32218; MUID:89098893; PMID:2563162  
 A.Molecule type: mRNA  
 A.Residues: 9-1153 <HIC>  
 A.Cross-references: GB:J04145; NID:9189068; PIDN:AAA59903.1; PID:9386975  
 A.Note: part of this sequence was confirmed by protein sequencing  
 R.Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.

J. Immunol. 150, 480-490, 1993  
 A>Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1 during evolution.  
 A.Reference number: A46526; MUID:93123748; PMID:8419480  
 A.Accession: A46526  
 A>Status: not compared with conceptual translation  
 A.Molecule type: DNA  
 A.Residues: 1-499, 501-1153 <FLR>  
 A.Cross-references: GB:S52227; NID:9263047; PIDN:AA24821.1; PID:9263049  
 A.Note: the last three bases of intron 13, CAG, are included in some but not all mature transcripts.  
 A.Note: sequence extracted from NCBI backbone (NCBI:121963)  
 R.Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.  
 Biochem. Biophys. Acta 874, 368-371, 1986  
 A>Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across species  
 A.Reference number: A90664; MUID:87076671; PMID:3539202  
 A.Molecule type: protein  
 A.Accession: A26091  
 A.Residues: 17-31 <PIR>  
 A.Experimental source: granulocytes  
 R.Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.  
 Blood 79, 865-870, 1992  
 A>Title: Characterization of the myeloid-specific CD11b promoter.  
 A.Reference number: 152567; MUID:92144986; PMID:1346576  
 A.Accession: 152567  
 A>Status: translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-9 <RES>  
 A.Cross-references: GB:M64477; NID:9180184; PIDN:AAA51960.1; PID:9553219  
 C.Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1  
 C.Genetics:  
 A.Gene: GDB:ITGAM; CR3A  
 A.Cross-references: GDB:120599; OMIM:120980  
 A.Map position: 16p11.2-16p11.2  
 A.Note: promoter contains a GATA motif and two Sp1 consensus binding sites  
 C.Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo  
 C.Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag  
 F.11-16/Domain: signal sequence #status predicted <SIG>  
 F.17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>  
 F.17-1108/Domain: extracellular #status predicted <EXT>  
 F.148-318/Domain: von Willebrand factor type A repeat homology <WMA2>  
 F.465-473/Region: calcium/magnesium binding #status predicted  
 F.530-538/Region: calcium/magnesium binding #status predicted  
 F.593-601/Region: calcium/magnesium binding #status predicted  
 F.1109-1134/Domain: transmembrane #status predicted <TM>  
 F.1135-1153/Domain: intracellular #status predicted <INT>  
 F.86,240,393,469,693,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding  
 Query Match 17.8%; Score 1085.5; DB 1; Length 1153;  
 Best Local Similarity 29.6%; Pred. No. 1.2e-72;  
 Matches 366; Conservative 196; Mismatches 492; Indels 183; Gaps 47;  
 Qy 11 LPVFLTC--LCSPFNIDENHRLPFPPEAEKRGVSIVQHGCGORMMLVGAPMDGPGD 68  
 Db 3 LRVLLTALTLTCHGFNDLTENAMTFQENARG-FQGSVQVQSGSR--VVGAPQETIVAN 58  
 Qy 69 RRGVYCPVCGANAPCAKGLDLYQGNSSHP-----ANVMHGMSLLETDPGG 120  
 Db 59 QRGSLYQC-----DYSTG-SCEPIRLQVPEAVMMSIGLSLAATSPQ 101  
 Qy 121 FMACAPLWSRACGSSVSSGICARVDAFPQGS LAPAOR-CPY-MDVIVLDGNSI 178  
 Db 102 LLACGPVHQCSBNYVKGICPLFGSNLRQPPQKFPBALRGCEQEDBDLFDGSGSI 161  
 Qy 179 YPMSEVQTLRLVGLKFLID-FQIQ-----VALVOGESPVHMSIGDRTAEVVRA 231  
 Db 162 IPHD-----FPRM--KEFVSTVMQKLKSKTLFSLIMQYSEFRITFKKEQNNPNPSL 214  
 Qy 232 AKNLSREGRETKTAQAIMVACTGFSQSHGSGREARLAVVVDGSSDOEEPLPAALKA 291  
 Db 215 VKPITQLGR-THATGIRKVRVRLFNITGARAKNAKILVITDGKEF-GDPLGYEDVI 272  
 Qy 292 CEAGR--VTRYGLAVLGHVLRQRPDPSSFLREITIASDPERFVNVTDAALTDIVDA 349

```

Db      273 PEADREGVIRYIGV-GDAFRSEKS-----ROELINTIASKPRPDHVFQVNNFALKTIONQ 327
Qy      350 LGDRIFGLESHAENESSFLEMSOIGFSTHRLKOGILFGMGAYDMGGSVLMLEGHRL 409
Db      328 LBEKFAITGTOTGSSSEFHEKSGGFAAITSNPLSTYGSDYMACGVFLYLSKES 387
Qy      410 PPRMALDEFPFALONHAAYLYGYSVSMILRGRRFLFSGAPRRFRHKVIAFOLKKG 469
Db      388 TFINMTRVDS-----DMNDAYIGYA-AAILNRNVGSLVGAIRYCHILVAMFR-QNTG 440
Qy      470 AVRVAOSLOGEOIGSFSELCPLDTRDGTDLVLAAPMELGPONKTKGVYVYL- 527
Db      441 MWESNANVKGTOIGAFGASLCSVDNSGSDVLIGAPHYY--EOTRGGVSVCPLE 498
Qy      528 GOOSLLTLCGTLOPEPPOD--ARFGFAMALPDLNOCGFADVAVGALPEDHGOALYLYG 586
Db      499 GGRARKOCCAVLYGEGQGWGRFGAALTVLGDVNGDKLTDVAIGABGEDNNGAYVLFPG 558
Qy      587 TG-SGVPRPAPORIAAASMPHALSYFGRSVDSRLDDGDLVAVAGAAIILSSRPI 645
Db      559 TSGSGISPSHSQRIAGSKLSPRLQYGGSLSGQDITMGVLDITVGAQHVLILRSQPV 618
Qy      646 VHLTSLSEVTPQALISVQORDCR--RGEAVCLTALCFQYTSRTPRGWDH---QFTWR 699
Db      619 LRVKAIMENPREVARNVEECNDQVYKKEAG--EYRVCLHYOKSTRRLREGOIQSVVT 676
Qy      700 FPAASLDEWTAGARAFDGSGRSLSPRLLSVG--NTVCQELPHVLD--TSDVLRPALTV 757
Db      677 YLALDSGRPHSRAPFNET--KNSTRQTOVIGLDTOTCTLKLQLPNCIEDVSPVLRL 734
Qy      758 TFLALNTTKPG-----PVINEGSPSIOKLVPFSKDCGDBECVTDVLQ---VNMDIRG 809
Db      735 NSLSVGTPLSAPGNLRPLVAEDAQRLFTALPFREKKGANDICQDDLSITFSFMSDC-- 792
Qy      810 SKKAPVVRGGRKKVLVSTTLNKRKNANVTLSISITFNSNL--HLASTPQR-ESPILV 865
Db      793 -----LVVGGPREFNVTYVRNDEDSYRTQVTFEPFLDLSYRKVSTLQNRORSQSMEL 846
Qy      866 ECAPSA-----HARLCSGVHPVQTCATVTFLEFFSCSSLSQVFKLTASSDRL 918
Db      847 ACESASSTSVGALSTSCSINHPIFPENSEVTFNITFVDSKASIG--NKL-----L 898
Qy      919 ERNGTLQENTAOJ-----SAYIOYEPH-----LFSSESTLHREYVHPYGT 959
Db      899 KANVTSSENMMPRTNKTPEQLPVKYAVVMVVTSHGVSTKYLNFTASEVTSVMQOY-- 956
Qy      960 LPEVGPEPEKTLRQNLGCIYVSGIITISA--LLPAVHAGNYFLSLSVT----- 1008
Db      957 -----QVSNLG---QRLPISLVFLVP-----VRLNQTVMDBRPQVTF 991
Qy      1009 TNNASCIQVNLTEPRPVPVPELOHTNRLNGSNTQCQVVRCHLQGLAKGTESVGLRL 1068
Db      992 SENLSTGCTTKERLBSHSPFLAEKRAPVNGSIAVCORIQCDIPFGIOEFPNATLKN 1051
Qy      1069 VANEFFRRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLLEV-VQTRPILISLWLI 1127
Db      1052 LSFDMYIKTSHNHLITVSTABILLPNDVSFTLLPGQCAFVRSQETKVEBPVNPPLIV 1111
Qy      1128 GSVLGLLLALLVFLCMLKGFPAHAKKIEBEKREK 1164
Db      1112 GSSVGLLLALLVFLCMLKGF-----KROYK 1139

```

## RESULT 9

cell surface glycoprotein CD11c precursor - human  
 N:Alternate names: leukocyte adhesion receptor p150.95 alpha chain  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 22-Jun-1999  
 C:Accession: A35584; A35543; S00864  
 R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
 J. Biol. Chem. 265, 12750-12751, 1990  
 A:Reference number: A35584

```

A:Contents: erratum
A:Accession: A35584
A:Molecule type: DNA
A:Residues: 1-1163 <COR>
A>Note: this revision to the sequence from reference A35543 includes the carboxyl end
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.
J. Biol. Chem. 265, 2782-2788, 1990
A:Title: Genomic structure of an integrin alpha subunit, the leukocyte p150.95 molecule.
A:Reference number: A35543; MUID:90153906; PMID:2303426
A:Accession: A35543
A:Molecule type: DNA
A:Residues: 1-834 <CO2>
A>Note: this sequence has been revised in reference A35584
R:Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.
EMBO J. 6, 4023-4028, 1987
A:Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte
A:Reference number: S00864; MUID:86166645; PMID:3327687
A:Accession: S00864
A:Molecule type: mRNA
A:Residues: 1-755, 'L', 757-1163 <CO3>
A:Cross-references: GB:M61695; EMBL:Y00093; NID:9487829; PIDN:AA59180.1; PID:9487830
A>Note: part of this sequence was confirmed by protein sequencing
C:Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on my
C:Genetics:
A:Gene: GDB:ITGAX; CD11C
A:Cross-references: GDB:119758; OMIM:151510
A:Map position: 16p11.2-16p11.2
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C:Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;
F:1-19/DNA: signal sequence #status predicted <SIG>
F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAY>
F:120-1167/DNA: extracellular #status predicted <EXT>
F:1149-319/DNA: von Willebrand factor type A repeat homology <VMA4>
F:1108-1133/DNA: transmembrane #status predicted <TM>
F:1134-1163/DNA: intracellular #status predicted <INT>
F:61,89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status pre

```

## Query Match 17.3%; Score 1054; DB 1; Length 1163;

Best Local Similarity 26.5%; Pred. No. 2.9e-70;  
 Matches 356; Conservative 201; Mismatches 474; Indels 220; Gaps 50;

```

Qy      13 LVFLTGLCSP--FNIDEHPRLPFGPPPAEFYGLVQHGSGQRMVLGAPWDCGSGRR 70
Db      8 LLLFTALATSLGFINLDELTAFL-RVDSAGCGDSVQVANS---VVVVGAPQKTTAAQOT 63
Qy      71 GDVYRCFVGAHNPACAGHIGDVLQNSHP-ANMHLGMSLLETDDSGFMACAPLWS 129
Db      64 GGLYQC--GYSTGACE-----PIGLQVPEAVNMSGLSLASTSPSOLLACGPVH 113
Qy      130 RACGSSVSSGICARVADSFOGSLAFTAORCPY--MDVIVLVDSNSIYP--WSEVQT 186
Db      114 HECGNMVLTLGCLFLGPT-QLTORLPVSRQECROBEDVIFLLDSSGSSSRNPATWMN 172
Qy      187 FLRLVGLFTDPEQIQLVQVGBSPVHEWSLGFRTKEBENVVRAAKLSRREGRETITA 246
Db      173 FVRAVISO-FQRP-STQPSLMQFNKQTHFTFEFFRTSNPLSLASVHQLQG--FTYTA 229
Qy      247 QAIWACTGESSGSGRPEARLLVVVTDGESH-DGEELPAALKACAGVTRGIAVL 305
Db      230 TAIQNVHRLPHASGARDATKILIVITDGKKGDSLDYQDVIPMADAAGIRIAYIVG 289
Qy      306 GHYLRQRDPSSFLREITTIASDPPERFFVNTDEAALTIDVALGRIIFGLESHAENE 365
Db      290 LAFQNR-----NSKMLEINDISKSGQEHIFVDEPDAKDIONQKKEIFALIEGTETSS 344
Qy      366 SSFGLMSQIGFSTHRLKOGILFGMGAYDMGGSVLMLEGHRLFPFPMALDEFPPLQ 425
Db      345 SSFELEMQEGFSVFTPDGCVLGAVGSFTV-----SGAFIYPPPMNS-----PFI 391
Qy      426 NHA-----AYIGYSVSMILRGRRFLFSGAPRRFRHKVIAF-OLKKGAVRVAQS 476
Db      392 NMSQENVMDRDSYLGYSLETALMKGVQSLV-LGAPRYOHTGKAVIFVOVSROW--RMKAE 448

```

```

QY 477 LQGEQIGSYFSGELCPDLTDGDTVDVLLVAAPMLGPNKKTGRVYV----- 524
D 449 VTGTIGSYFSGELCPDLTDGDTVDVLLVAAPMLGPNKKTGRVYV----- 506
QY 525 --YVGGQSLTLTGLOPEPPDARFPGAMGALDLDNDGADVAVAGAPLEDHOGALY 582
D 507 DAVLVGEQG-----HPW-----GRFGALITVLGVNDGDKLTDVIGAPGSEENRGA 554
QY 583 LVHGQTQS-CVRPAPQRIAAAMPALSYFGSVDRGLDLDDGDDVDVAVAGQAAILLS 641
D 555 LHHGVLPSPISPSHRSRIAGSLSRLQYFGALSGGDLTDGDLVAVAGRGVLLIR 614
QY 642 SRPIYHLTPSLEVTPOAISVQDCRRR-GQAVCLTALCFQVTSRTP--GRWDHGFY 697
D 615 TRPVLVWVSGQFIAPAEIPRSAFECREQVSEQLTVQSNICLYIDKRSKNLGRDLQSS 674
QY 698 MAFIASLDWTAGAPAFDGGQRLSPRLRLSVGNVTCQQLHFVLDTS--DYLPRPA 754
D 675 VTLDLALDGRSLSPATFQETKRR-SLSRVRLGKAKCE--NFMLLPSCYEDSVTPIT 731
QY 755 LVTVFALDNTTKP-----GVLNEGSPTSIOKLVPFGKDCGPNCEVTDVLQVNDI 807
D 732 LALNFTL--VGKPLAFRLRLPMLAADQRYFTASLPFKKCGADHICQDNLGIFSFP- 788
QY 808 RGSRAKPVVGRGRRKVLVSTLE-----NRKEMAVNTSLSTIFS----- 847
D 789 -----GLKSLVGSNLEINAEVWVNDSDSYGT--TTFSPAGLSYRYVAE 834
QY 848 -----RNLHLASLTPQRESPIKVECAAPSAHAALCSGHPVFPOTGAKVTLLEFSPC 900
D 835 GQKQGLSLHLTC---DSAPV---GSGGTWSTSCRINHLIFPGAGITFLATFVDS 886
QY 901 SSLLSQVFGKLTASSDSLEKNGTLOENTAO-----TSAYIQVEPHLLS--- 944
D 887 KAVLDRL-LLTANVSENNTPTRTSKTTQLPKVYAYTVVSSHQETKTLNSESSE 945
QY 945 --SESTLHREYVHPYG--TLPVGPGEFETTLRVONLGYVSGILISALL--PAVAGG 998
D 946 KESHVAMHRVYQVNNIGORLPLV-----SINFWPEVLEINQEAVMMDVEVSHQ 992
QY 999 NYFSLSQVITNNACIYONLTERPGRPVHPELQHTNR--LNSNTQCVVRRCHLQ 1055
D 993 NPSLCS-----SEKLAIPA-SDFLHIOQNPVLDSIAGCLRFRCVDSF 1037
QY 1056 AKGTESVGLRLVHNEFFRAKFKSLTVSTFELGTEBGSVLQTEASRMSSESLLEVQ 1115
D 1038 SVQEELDPLTKNLSFGWKQQLQKKVSVSAEITFDTSVYSQLPQGAFFRA-----Q 1092
QY 1116 TRPILISMI-----LIGSVLGGILLALLLVFCIMKLGFF--AHKQIPPE 1158
D 1093 TTVLEKKVHNPTPLIVGSSIGILLALLITAVLYKVGFFRKQYKEMEE 1143
Db
RESULT 10
S00551
leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse
N:Alternate names: complement-3 receptor alpha chain
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1989 #sequence-revision 30-Sep-1991 #text-change 22-Oct-1999
C:Accession: S00551; 159078
R:Pyteia, R.
EMBO J. 7, 1371-1378, 1988
A:Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the
A:Reference number: S00551; MUID:862812584; PMID:3044779
A:Accession: S00551
A:Molecule type: DNA
A:Residues: 1-1153 <PVT>
A:Cross-references: EMBL:X07640; NID:952982; PID:CA0479.1; PID:952983
A>Note: the authors translated the codon CAC for residue 569 as Gln
R:Sastre, L.; Roman, J.M.; Teplow, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Roberts,
Proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986
A:Title: A partial genomic DNA clone for the alpha subunit of the mouse complement recep
A:Reference number: I59078; MUID:86287312; PMID:2942940

```

```

A:Accession: I59078
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 11-44 <RES>
A:Cross-references: GB:M14293; NID:g198993; PID:AAA39484.1; PID:g554193
C:Genetics:
A:Gene: Mac-1
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C:Keywords: cell adhesion; glycoprotein; transmembrane protein
F:1-16/Domain: signal sequence #status predicted <SIG>
F:117-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental
F:1148-318/Domain: von Willebrand factor type A repeat homology <VMA2>
F:1106-1129/Domain: transmembrane #status predicted <TM>
Query Match 16.7%; Score 1022.5; DB 2; Length 1153;
Best Local Similarity 28.6%; Pred. No. 6,66-68;
Matches 353; Conservative 218; Mismatches 481; Indels 181; Gaps 50;
QY 13 LVPLTGLCSFNLDEHHRLEPPPEAEFGYSVLQHVGGGRMLVGA PMDPSGDRRD 72
D 7 LVTLALCHGNLDTHEHMTFQENAKG-FQGNVQ-LGTS--VVAAPQAKAVNQGA 62
QY 73 YVRCPVGAHNAAPCAKHLGDIQGNSSHP-----AVNMHLGMSLLEFDGCGFMAC 124
D 63 LYQC-----DYST-SRCHPIPLQVPEPAVMSGLSLAVSTVPQQLAC 105
QY 125 APLMSRACSSVFSSGICARVDAS-FQPGSLAPTAORCPY--MDVVLVGSNSI--YP 180
D 106 GPTAHQNKENTYVNGLCYLFSGNULRPPOGPPEALBECPOQESDVIYLIQSGSINNID 165
QY 181 WSEVQTLRLRLVGLFLIDPEQIQ-----VGLVQGESFVHMSLGFRTKEEVRAAKNL 235
D 166 FQKMEFVSTW-----EQFKSKTLFSLMQSDERFIHFTFNDPFRKNSPSRSHVPI 218
QY 236 SRREGREKTKQALMVACTEFGSOSHGCRPAAALLVYVTDGSHDEELPAALKACAG 295
D 219 KQLNGR--IKTASGARKVRELFLHKTNGARENAAILTVITDGKRF--GPDLYKQVIRPAD 276
QY 296 R--VTRVGIAYLGHYLRQRDPSSFLREIRITASDPDERFPFNVTDEALTDIYDALGDR 353
D 277 RAGVIRYIVGY-GNAFNK--PQS--RELDITAKRPAHEHVFQVDNFEALMTIONQLEK 331
QY 354 IFGLEGSHAENESSFGELMSQIGFSTRKLDGILFGWGVAYDM--GGSVLMLEGHRLEPP 412
D 332 IFALEGTGTSTSEBHEMSQEGFSASITSNGLIGVSGSDWAGAFLYTSKQVPTIN 391
QY 413 RMALDEPRLALONHAAYLYGSVSSMLIRGGRPLFLSGAPFRFRKVIYALQKKQAVR 472
D 392 TTRVDSDM-----NDAYLGVA-SAVILRNVOQLVLGAPRYOHIGLVWFR--ENFGTWE 443
QY 473 VAOSLQGEQIGSYFSGELCPDLTDGDTVDVLLVAAPMLGPNKKTGRVYVLYGQ--- 529
D 444 PHTISQGISYFSGELCSVDMDADGNTNLTILGAPHYV--EKTRGGQVSCPLPRRA 501
QY 530 --QSLLTLOGLTQPEPPDARFPGAMGALDLDNDGADVAVAGAPLEDHOGALYLYHG 587
D 502 RMQGEALLHGP-QGHR--WGRFGALITVLGVNDGDKLTDVIGAPGEGENGAAYITFGA 558
QY 588 Q-SGVRRPAPQRIAAAMPALSYFGSVDRGLDLDDGDDVDVAVAGQAAILLSRPIV 646
D 559 SIASISASHSHRIIGAHSPGLQYFGSLSGGKDLTMDGMLDVLVAGQGHILLRAQPV 618
QY 647 HLTPSLEVTPOAISVQDCRRR-----GQAVCLTALCFQVTSRTPR---WDPQF 696
D 619 RLEATMESPKKVARSPACQOVLKNDAGEVYCL-----RVKNTKRLREBGLIQS 672
QY 697 YMRFTASLDWTAGAPAFDGGQRLSPRLRLSVGNVTCQQLHFVLD--TSYLRLPYAL 755
D 673 TVTVDLALDVPKRSIRAFEDET-KNNTRRRQVGVLMQKCTKTLILIPDCVDSVSPITL 731
QY 756 TVTVALDNTTKP-----GVLNEGSPTSIOKLVPFGKDCGPNCEVTDVLQVNDIR 808
D 732 RLNYTL--VGKPLRSFGNLRVLAMDQRYFTAFPEKNGCNDISICODD--LSITWSAM 787

```



158409  
 Integrin alpha-9 chain precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 31-Dec-2000  
 C:Accession: I58409; A49459  
 R:Hihi, K.; Yamakawa, K.; Ueda, R.; Horio, Y.; Murata, Y.; Tamari, M.; Uchida, K.; Takah  
 Onogene 9, 611-619, 1994  
 A:Title: Aberrant upregulation of a novel integrin alpha subunit gene at 3p21.3 in sma11  
 A:Reference number: I58409; MUID:94119603; PMID:8290272  
 A:Accession: I58409  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1035 <RES>  
 A:Cross-references: GB:D25303; NID:9464180; PID:BAA04984.1; PID:9533327  
 R:Palmer, E.L.; Ruegg, C.; Ferrando, R.; Pytel, R.; Sheppard, D.  
 J. Cell Biol. 123, 1289-1297, 1993  
 A:Title: Sequence and tissue distribution of the integrin alpha9 subunit, a novel patrine  
 A:Reference number: A49459; MUID:94064789; PMID:8245132  
 A:Accession: A49459  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 30-1035 <PAL>  
 A:Cross-references: GB:L24158  
 C:Superfamily: integrin alpha-4 chain  
 C:Keywords: glycoprotein; metal binding; transmembrane protein  
 F:1-27/Domain: signal sequence #status predicted <Sig>

Query Match 12.1%; Score 738; DB 2; Length 1035;  
 Best Local Similarity 22.6%; Pred. No. 1.3e-46;  
 Matches 284; Conservative 179; Mismatches 446; Indels 350; Gaps 44;

9 LFLPLVFLTGLCSFNLDEHNRPLFPGRPEAFEGSVLQHVGGGQRMVLGAPMD---G 64  
 16 LLLALVVGIPAGAVNLDPQRPVHFGPADSFQVAVLEHFDNTRMVLVGA PKADSKYS 75  
 65 PSGRGRGVYRCVPVGCANAP-----CAKGLADYOLGNS-SHPAVNMLGMSL-E 114  
 76 PSVKSPGAVFKCRV---HTNPDRCCTELDMARKRGKRGSCGKTCREDRDDEMGVSLARQ 132  
 115 TDGDDGFMACAPLWBRAC--GSSVFSGICARVADSPQGGSLATPAORCPVYMYVIVL 172  
 133 PKADGRVLLCAKRMKNTIYEAHDHLPFGCYIIPSNLQAKG----- 173  
 173 DGSNSIYVSEVQTFELRLVGLFIDPEQIQVGLVQVGGSPVHMSLGDPRTEKEVRAA 232  
 174 -----RT----- 175  
 233 KNLRSREGRETKTAQAINVACTEGFSQSHGGRPEARLLVVTDESHDEELPALKAC 292  
 176 -----LIPCYEEYKKY-----GEEHG-----SC 194  
 293 EAGRTYRGIAVLGHYLRQRDPSSFLKIRITIASDPDRFFFNVTDEALTDIYDALGD 352  
 195 QAG-----IA----- 199  
 353 RIFGLEGSHAENESSFGLMSQIGFSTRHLKDGILFGWAGVADMGGSVLMLEGRHLP 412  
 200 -----GFTTEL---VVMGAPGSYVMAGTIKVLNLTNTY-- 231  
 413 RMALEDEPPPALONTAAYIGYSVSSMLRGGRRL-FLSGAPRRFRHGRVLAFLQ-KKQGA 470  
 232 -LKLNDDEV--IMNRRTYLVGAVTAGHFSHPSTIDVVGAPQDKIGKVIYFRADRRSGT 288  
 471 VRVAOSLQGEQIGSYFGSGLCPDLDRTDGTDLVLLVAAPMFLGPKNGKEGRVYVVLVQ 530  
 289 LKIRIQASGKMGKSGYFGSSLCVNDLNGDGLD-LLVGAPMF--SIRREGQVTVYINRGN 345  
 531 SLTLTGTLQPPPPDARFGAFMGALPDLNQGFAVDVAVGAPLEJHGOGALVLYGTGSG 590  
 346 GALEEQALVLTGGAANAHGSESIALDLDNDGFFDVVAIGAPKEDDFAGAVIYHGDAG 405  
 591 VRHPDAQRIAAASMPHALSYFGRSYDGRLLDGDLDVVAVGA--QGAAILLSRPVHL 648

406 IVPOYSMKLSGQKINPVILRMFGQISGGIDMDNGYDPVTVGAFMSDSVLLRARPVITV 465  
 649 TPSLEVTPOAISVQDRRRRGQAVCTALACF-----QVTSRTP 689  
 466 DVST-FLPGSINITAPQCHDQOQEVNCLNVTTCSSFKHAYPEEIGLVYLMADYAKKX 524  
 690 GRMDHQFYMRFTASLDENTAGARAAFDGSGGRLSPRLRLSVGNVCEQLHFHV-LDTS 748  
 525 GQMPREVYVLLGETMGVQT-----EKLQTYMBETCHYVAHVAVRQVD 568  
 749 YLRPALVTAL-----DNTTKP-GRVL--NEGSPISIQKLVPPSKCGPDNECVT 797  
 569 VISPIVFEPAAYSLEBHVTEBERELPLTPVLRKKKGCKLAQKQVTFERNCRSE-DCAA 627  
 798 DLVQVNMNDIG-SKAPFVVRGGRKVLVSTLENKRNAYNLSLTIFFRNILHSLT 856  
 628 DLQAGKLLSSMEKTYLALGAVKINSLNISISNLDGDAYDANVSFNVSRELFFIMW 687  
 857 FQRESPIKECAAPSARHLCVGHVVFQTAGKTVFLLEFFSGSSLSIOVFGKL-TASS 915  
 688 QKEE--MGISCELLESDPLKCSGVFPFMRSKYEFVIFPTSHLSGEEVLSFIVTAQS 745  
 916 DSLERNGLQENTQTSAYIOYE-----PHLFSSEST-----LHRYEVRHG 958  
 746 GNTRESSELDHNTLVLMVPLMHEVDTSITGMSPTSFYVGSVDANFTQDDBLCH--- 802  
 959 TLVPQGPBEFTTLRVONLGCYVVSGLIISALP-AVAGGNYFLSLQVIT--NNASCI 1015  
 803 FQPI-----NITQVNTGPTSLPGSSVSISFNRRLSSGAEHFNQEMVVGQEKNGS 856  
 1016 VQNTPEPPRPVHBELOHT-----NRLNGSNTOCQVVRCHLGQLAGTEVS 1063  
 857 FQKPTPCIIQOGENIHFTIFAFFTKSGRAVLDCERKGISCLTAHGNFALAKEESRTI 916  
 1064 GLILVANE-----FFRAKFK--SLVYVSIFELGTEGVSYLQLTESRMS 1109  
 917 DIYMLNTEILIKKSSSVIQCMSRAKVKVDPALVEIAHGNPREVTV--VFEA----- 968  
 1110 LLEVQTRPILISWILIGSVLGLLALLVFCMLKGF--AHKKIPEEKKEBKLE 1166  
 969 LHLNBERGVVGG-WIATISLLVGLIFLLAVLMLKMGFFRRRYKEIIEAKRKENE 1025

RESULT 13  
 JC7294  
 alpha9 integrin - sea urchin (Strongylocentrotus purpuratus)  
 C:Species: Strongylocentrotus purpuratus (purple urchin)  
 C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: JC7294  
 R:Suban, J.M.; Just, M.L.; Lennarz, W.J.  
 Biochem. Biophys. Res. Commun. 272, 929-935, 2000  
 A:Title: Cloning and characterization of alpha9 integrin in embryos of the sea urchin St  
 A:Reference number: JC7294  
 A:Contents: Embryo  
 A:Accession: JC7294  
 A:Molecule type: mRNA  
 A:Residues: 1-1054 <SUS>  
 A:Cross-references: GB:AD55724  
 C:Genetics:  
 A:Gene: snalhap  
 C:Superfamily: integrin alpha-2b chain  
 C:Keywords: calcium binding; embryo; glycolysis; glycoprotein; heterodimer; transmembran

Query Match 12.0%; Score 732.5; DB 2; Length 1054;  
 Best Local Similarity 23.4%; Pred. No. 5.3e-46;  
 Matches 294; Conservative 188; Mismatches 432; Indels 345; Gaps 47;

10 LFLPLVFLTGL-----CSFNLDEHNRPLFPGRPEAFEGSVLQHVGGGQRMVLG---AP 61  
 4 FLLSFTCTVLVLDSTYAFNFDLRAVYKFDGPOSSLGFSYAOHRDONTDVTLGABAP 63  
 62 WDGPGRGRGVYRCVPVGG-AHNAFCAK---GHIGDYOLGNSSHPAVNMHLMGMSLBTGD 117

```

Db      64 TTQPGVINGAAYKCPVTPLSGSGPCEQVFPDITGTEVLDSK-----NQMFATLASGP 119
Qy      118 DGFMACAP--LMSRACGSSVFSSGICARVDASFQOGSLAPRACRPTMYDVIVLDS 175
Db      120 DGIILACAPRLVWLQT-----STISPT----- 141
Qy      176 NSIYPMSEVQTFRLRIYVKLFIIDPEQIOVGLVOYGESPVHEWELGDPRTKEEVRAAKUL 235
Db      142 ----- 141
Qy      236 SRREGRETYTAQAIWVACTEGFSQSHGGRPEARLLVVTGDSHDEELPALKACENG 295
Db      142 -----DKEREPT----- 148
Qy      296 RYTRGIVAVLGHYLRQRDPSSFLREIRTIASDPDERFFPNVTDEALDIYDALGRIF 355
Db      149 -----GTCFVGH-----SDFTNFVNSPCOSTDRD-----LF 175
Qy      356 GLEGSHAENESSFGLMSOIGFSTHRLKDG--ILFGMVAGAYDGGSVLWLEGHRLFPPR 413
Db      176 GFD-----KITHCQAGFSAQIPSDNSTLVMKAPSY-----YLOG--QIFAGS 216
Qy      414 MA-LED-----EPPALQNHAAVLYGVSSSMLLRG--GRRLFLSGARFRFH--RCKVIAPOL 465
Db      217 LSTLSDVSNTPEDQAVAFDN--SYRGYSALAGDFNGDGLDYYVGTFRGESLGLVAIFD- 273
Qy      466 KDGAVRVAQSLQGEQIGSYFSELCPLDTRDRCGTTDVLVVAPEML-----GQNKETGR 521
Db      274 --OSLVEIITPVVGEQIVSYFGSVASVDVNGDGLD--LLVGAPMTNREPAETKEWEAR 330
Qy      522 VVYVL-----VQOQSLLTLQGTLOPEPPODARFGFAMGALPDINODGFADVAVAGALE- 574
Db      331 VVYVLQNMADHSIGAPQMLTGKTI-----RAFGFPIISIGSNDGDNDAIGAPYVG 383
Qy      575 DGHQGLVLYHGTQSCVRPHAPRIAAASMPHA--LSYFGESVDGRILDDGDIYDVAVGA 633
Db      384 ENDSGVVYLYHGSABEGLRLTESQVLTLPSELGFSDITTFGFSVGGQDMQONDYDPLVGA 443
Qy      634 QG--AAIILSSRIYVLTSLSEVTPQAISVQGRDCR--RRQGEAVCLTALCFQVT--SRFP 689
Db      444 ESADAALIVTRRVLLLEBELTEPIGINDKNTYLPDGTMYTSVVAACFYTGNNLP 503
Qy      690 GRWDHOFYMRFTASLDEWTAGAARAFDSGQRLSPRLRLSVGNVTCLOLHFEHLDT--SD 748
Db      504 ARIGISYTLTVDSI--TSGRALLLEVNBLSQVTKNRMLDVGMRKCDPLRAVYVNTID 560
Qy      749 YLRPAVLYTFALDNTT-----KPGVLAEGSPTSIGKLVPSKCGPNECVTDLVLOY 803
Db      561 KLTPLIVDQYELTDESILLPEILPIINKEAVSQTQVSIQNNC--VNNICIEPILGIV 619
Qy      804 NMDIRSRKAPFVVRGRRKVLSTLENKENVYTSLSIIFSRMLHLASLTPORESPI 863
Db      620 TPNL-----PNIVIQAOBELTLVSIINNGEDAFQSTILAVVYPELQYVRLERRANMF 673
Qy      864 KYECAPASAHARL--CSVGHFVFQTAGKVTFLLEFFESCSLSLQVFGK-----LTASS 915
Db      674 SYTCTEDSALRMITCOTGNPLVG-----KYNLEFGILTSL--QVSGDMONIEFYLVAAS 726
Qy      916 DSELRNGTLOENTAOISAYIOYEP--HLFSSSESTHRYEAPHYGLPLVGP----- 964
Db      727 ENNEEDNNTLDNNEILANTVAIVDATLKLISASYPEIVTYRVPEDNIVPEFPKNSSEAD 786
Qy      965 GPEFKTTLVONLGCVVVSGLIISALLPVAHGNYFELSIOVITN--NASC--IVONLTPR 1022
Db      787 GHEVHLHYEVRNTGSSNAAEVTLINRWPEKENGDIYLLGLIMDEGVTCQISQOQANP 846
Qy      1023 PGPVPAPEELOHTNRLNGSNTQ-----COVVRCHIGQL 1055
Db      847 LGVKL---EASTKEQISNSTTQVSGRRKREGEVAAEALQAAPIFCPPECVCLINCTIDBI 903
Qy      1056 AKTEVSVGLRLVHNEFFRRRAKFKSLTVVSTFELTEGGSV-----LQLTASR 1105
Db      904 KATKSKVIRILGFRFERTOKAVSEAVPIYOVLASTATATVRSIPYNIPLIMEFTDSTK 963

```

```

Qy      1106 WSESL-LEVQTRPRLISLWILIGSLGLLLALVFCWKLGFFAKKTIPEEKRE 1163
Db      964 ASTLTABELVPPVSIAMWIIIVSVLGGIILLIILIGLKKCGFFERKKEGDEQKEYE 1022

RESULT 14
A:1131
lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse
M:alternate names: Integrin alpha-4
C:Species: Mus musculus (house mouse)
C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 24-Sep-1999
C:Accession: A41131; S16742
R:Neuhaus, H.; Hu, M.C.T.; Hemler, M.E.; Takada, Y.; Holzman, B.; Weissman, I.L.
J. Cell Biol. 115, 1149-1158, 1991
M:Title: Cloning and expression of cDNAs for the alpha subunit of the murine lymphocyte-
A:Reference number: A41131; MUID:92064645; PMID:1840602
A:Accession: A41131
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1039 <NEU>
M:Cross-references: EMBL:X53176; NID:951484; PIDN:CAA37316.1; PID:951485
C:Superfamily: Integrin alpha-4 chain
C:Keywords: cytoskeleton; transmembrane protein

Query Match
Best Local Similarity 22.7%; Pred. No. 5e-43;
Matches 286; Conservative 186; Mismatches 396; Indels 392; Gaps 59;

Qy      23 FNLDHNPRLPPGPEAEFGYSVLQHVGGQRMVLVAP--W--DGPSGDRRGDYRCPV 78
Db      41 YLIDENALLVQGPBGTLFGYSVLHSHGSRKMLIVGAPTASMLSNASVNPALYRCGI 100
Qy      79 -----GGAHAPCAKGHLDYQULGNSHPAVNMHLGMSLLETGQD--CGFMAC 124
Db      101 RKNPNQCEQLOSGSPGCEPGKTCLEERD-----NOMLVLTLSRQENGSIYTC 151
Qy      125 APLMRACGSSVFSFGICARVDASFQOGSLAPRACRPTMYDVIVLVDGNSIYPMSEV 184
Db      152 GHRW-----KNIF-----YMKSDNKLPT----- 169
Qy      165 OTFLRLVGLKFLIDPEQIOVGLVOYGESPVHEWELGDPRTKEEVRAAKULSRREGRETK 244
Db      170 -----GICVMP-----SDLRTE-----LSKR----- 186
Qy      245 TQAQIMVACTEGFSQSHGGRPEARLLVVTGDSHDEELPALKACEAGRVTRYGIAY 304
Db      187 -----MAFCYKDYTRKFGEN-----FASCOAG----- 208
Qy      305 LGHYLRQRDPSSFLREIRTIASDPDER-----FFPNVTDEAALTDIVDALGRIFGLE 358
Db      209 -----ISSFTQDOLIMGAPSSYWTGYTFVNYIT-----TNOYKAFVDR----- 248
Qy      359 GSHAENESSFGLMSOIGFSTHRLKDGILFGMVAGAYDGGSVLWLEGHRLFPPRALED 418
Db      249 -----QNVKFG----- 255
Qy      419 EPPALQNHAAVLYGVSSSMLLRGRRL--FLSGAPRFRHGRKVIAPOLKDGAVRVAQSL 477
Db      256 -----SYLGSVAGHFRSPHTTEVVGGAQHEQIGKAYIFSI--DEBELIYEM 304
Qy      478 QGEQIGSYFSELCPLDTRDRCGTTDVLVVAPEMLGPON--KETGRVVYVVLVGQOS--LL 533
Db      305 KKKKIGSIFGASVCAVDADNADGFD--LLVGAPEM-----OSTIREGRVYVYINSGGAMV 359
Qy      534 TLQGTLOPEPPODARFGFAMGALPDINODGFADVAVAGALEDGHQGLVLYHGTQSCVPR 593
Db      360 EMERVLVGSDDKYARFSGESIANLGDIDNDGFEDIAIGAPQDDIDRGAUYIYNGRADISS 419
Qy      594 HPAQRIIAASMPHALSYGRSVDRGLDLDGDDIYDVAVGA--QGAAILSSRPVYHLTPS 651
Db      420 TYSQRIEQQQSLSLRMGQSIGQSIDADNNGYDVAVAGAFQSDSAVLLRTRPVIVVAS 479

```



```

QY 652 LEVTPQAISVVDRCRRRQGEAVCLTALCFQVTSR-TPGRMDHQFYAFSTASDE----- 706
Db 480 IS-HBESVRIKFDCTENELPSVCHMLTLTLCFSYKGEVPG-----YIVLFYNNLSLVHRA 534
QY 707 -----WTAGARAAFDGGGRLSPRLRLSVGNVTCGLH--FHVLDTSYLPAVLT 756
Db 535 ESPSRFYFSNGTSDVITGS-----IRVSSSGEKC-RTHQAFMRKDVRDILTFPIHVE 585
QY 757 VTFALDN--TTKPG-----FVINEGSPTS-IOELVPSKDCGPDNCEVTLVLOVM 805
Db 586 ATYHGHVITKRNTEEPPLQPILOQKKEXDIRMINFARFCAVEN-CSADLQVAKV 644
QY 806 D-IRSGRAKPFVVRGGRKVLVSTLENRKENAVNTSLITSRLMLASLTPOGESPJK 864
Db 645 GFLKPEYNTYLAVGSMTKIMLVSLFNAGDDAYETTLNVQLPTGLYFPIKILDEBKQIN 704
QY 865 VECAPSAHARL-CSVGHVFPQTGA--VTFPLLEFEFSSCSLLSQVFGKLTAS-SDSLER 920
Db 705 CEVTSSSGIVKLACSLIGIYVDRSLRIDISFLDV-----SLSRAHEDLSIVHASCE 759
QY 921 NGTL---QENTAQTSAYIQEYEPHLLFSSSESTLHRYEHP---YGTLPVGPGE----- 967
Db 760 BEELQVNRNRVTLTLPLEVEML-----TVHGL-VNPTSPVYGSSEENE-PETCMAEK 811
QY 968 FTTLRVQNLGCVVSGLIISALLPAVHAGNYFL-----SLSQVITNNASC----- 1014
Db 812 LNLTFHVINTGISMAPNVSVKIMVP-----NSFLPQDDKLFNVLDVQTTGQCFHKHYG 865
QY 1015 -----IYQNLTEPPGPVHPELOHTNR-----LNGSNTQCVQVRCHIGOLAKGTE 1060
Db 866 RECTPAQOKGIAGTLTD-----IYFELSKTDKRLLYCMKADQHCIDPLCNGKMSGKE 919
QY 1061 VSVGLRLVHNEFFRRAKFSKSLTVVST--FELGT-----EEGSLVQLTFASRWSLSLE- 1112
Db 920 AS-----VHIOLEGRPSILEMDESSLSLKEIKATAFPPHPKVIENKDEVAVHFLDG 973
QY 1113 VVQTP-----TILSLMLIGSVLGGLLALLVLC-LMKLGF--AHKTIPEEKRE 1163
Db 974 LHHQRPKHFTIIITISILLG-----LIVLLLSICVMWKAGFPKQKSILOEENRD 1027

RESULT 15
S06046
integrin alpha-4 chain precursor - human
M:Alternate names: CD39d; very late antigen-4 alpha chain; VLA-4 alpha chain
C:Species: Homo sapiens (man)
C:Date: 01-Dec-1989 #sequence revision 01-Dec-1989 #text change 24-Sep-1999
C:Accession: S06046; A39355; D28018
R:Takada, Y.; Ellices, M.U.; Crouse, C.; Hemler, M.E.
EMBO J. 8, 1361-1368, 1989
A:Title: The primary structure of the alpha(4) subunit of VLA-4: homology to other integ
A:Reference number: S06046; MUID:89356603; PMID:2788572
A:Accession: S06046
A:Molecule type: mRNA
A:Residues: 1-1038 <TK>
A:Cross-references: GB:X16983; EMBL:X15356; NID:G33945; PIDN:CAA34852.1; PID:G33946
R:Rosen, G.D.; Birkenmeier, T.M.; Dean, D.C.
Proc. Natl. Acad. Sci. U.S.A. 86, 4094-4098, 1989
A:Title: Characterization of the alpha-4 integrin gene promoter.
A:Reference number: A39355; MUID:91239513; PMID:2034655
A:Accession: A39355
A:Molecule type: DNA
A:Residues: 1-93 <RNA>
A:Cross-references: GB:M62841
R:Takada, Y.; Strominger, J.L.; Hemler, M.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 3339-3243, 1987
A:Title: The very late antigen family of heterodimers is part of a superfamily of molecu
A:Reference number: A94151; MUID:87204112; PMID:3033641
A:Accession: D28018
A:Molecule type: protein
A:Residues: 40-50, 'E', 52-53 <TA>
C:Genetics:
A:Gene: GDB:ITGA4; CD49D

```

```

A:Cross-references: GDB:128032; OMIM:192975
A:Map position: 2q31-2q32
C:Superfamily: integrin alpha-4 chain
C:Keywords: cell adhesion; cytoskeleton; duplication; glycoprotein; heterodimer; transme
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-1038/Product: integrin VLA-4 alpha-4 chain #status predicted <MAT>

Query Match          11.2% Score 684; DB 2; Length 1038;
Best Local Similarity 22.9%; Pred. No. 1.4e-42;
Matches 289; Conservative 181; Mismatches 423; Indels 368; Gaps 54;

QY 9 LFLPLVLTGLCSPPNDENHRLPPRPAEFGYSVLQHYGGQRMVLVGP--W-DG 64
Db 28 LLLCLGVPTG--RPYNDVESALLYQGPHNTLFGSVVLSHGANKRMVLVGAPTANMLAN 85
QY 65 PSGRGRGVNRCPVG-----GANHA-PCAKGLGDYOLGNSHPANMHGLMS 111
Db 86 ASVINPQAIYRCRGNKPNQGTCELOLQGSFNGEPCGKTCLEERD-----NOMLGV 136
QY 112 LLETQD--GCFMACAPLWSRACSSVSSGICAEVDASFOQSLAPTRQCPYMDVYI 170
Db 137 LSRQGENGSIVTCGHR----- 154
QY 171 VLDSGNSLYPMSEVQTFRLRLVGLFLIDPEQIQVGLVOYGESPVHMSLGPFRTKEVVR 230
Db 155 ---KNIFYINE-----NKLPTGGC-YGVPP-----DLRT----- 180
QY 231 AAKSLRSREGRETTQAQIMVACTEGFSQSHGCRPEARALLVVTDSHGEEELPALK 290
Db 181 -----ELSKRIA 187
QY 291 ACEAGRVYRGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFFNVTDEAALTDIYDAL 350
Db 188 PCYQDYVKKF----- 198
QY 351 GDRIFGLEGSNAENESSGLEMSQIGFSTHRLKQGLFGMVGAYDWGSYMLF---CGH 407
Db 199 -----ENFAS-----COAGISSFTKDLIWGAGSSYWTGSLFVNYITNKY 241
QY 408 RLPPRMALEDPEPPALONH---AAYLGYSVSMMLGGRRL-PLSGAPRRRHGKVIAF 463
Db 242 KAFIDK-----ONOVKFGSYLGYSVGAQHFSQHTTEVAVGAPQHEQIGKAYIF 290
QY 464 QLKKGAVRVAQSLQGEIGSYFSGSELCPDITDRDGTITDVLVLAAPMELGPON--KETGR 521
Db 291 SI-DEKELNIIHEMKKGLSGYFGASCAVDLNDGSSD-LLVGAPM---QSTIRREGR 344
QY 522 VYVYL-VGQSL--TLQGTLOPPPODARFGFAMGALPDLNQGPADYAVGAPLEDGHQ 579
Db 345 VFVYINSGSGAVNMAMETNLVGSDKYARFGEISIVNLGDIINDGFEDEVAIGAPOEDDLOG 404
QY 580 ALVYHQSQSVRRHPHQRILAAAMPHALSYFGGSVQGRDLDDDDLVDVAVGA--QGA 637
Db 405 ALIYINGADDISSTFSQRIEGLQISLSLWFGGSIQGLDADNNGYDVAVGAFRSDA 464
QY 638 ILLSRPPIVHTLPSLEVPQASIVVORDCRRRQGEAVCLTALCFQVTSR-TPGRMDHQF 696
Db 465 VLLKTRPVVLYDALLS-HBESVNRKTFDVCVNGMPSCIDILTLFSYKGEVPG---YI 519
QY 697 YMRPTASID--EWTAGARAAFDSSG--QRLSPRLRLSVGNVTCGLH--FHVLDTSYL 750
Db 520 VLFYMSLDVNRKAESPFRFYSSNGTSDVITGSIQVSSRANC-RTHQAFMRKDVRDIL 578
QY 751 RPAVLTVFL-----DNTTKGPVIANESSPSI--OKLVFSDDCGPDNCEVTL 799
Db 579 TPIQIEAAVYHGHVVISKSTEEPPLQPILOQKKEXDIRMINFARFCAHEN-CSADL 637
QY 800 VLQVMD-IRSGRAKPFVVRGGRKVLVSTLENRKENAVNTSLITSRLMLASLTPOG 858
Db 638 QVSAKIGFLKHEKHTYLAVGSMTKIMLVSLFNAGDDAYETTLNVQLPGLYFPIKIL 697
QY 859 RESPIKVECAPSAHARL-CSVGHVFPQ--TGAKVTFLEFEFSSCSLLSQVFGKLT-- 912

```

```
Db      698 EERQINCEVTNDSGVVQDSCISGYIVVDHLSRIDISFLLDV-----SSLRAEEDLSITV 752
QY      913 -ASSDSELRNGTLQENTAGTSAVIOYEPHLLPSSSESTLHREYHP-----YGTLPVGPGE 967
Db      753 HATCENEEMDNKHSRVTVAIPKYEKL-----TVHGF-VNPTSFVGSNDENE-PE 804
QY      968 -----FKTLRVONIGCYVVSGLIISALLP-AVAHGNYFLSLSCVITNNASCIQONLT 1020
Db      805 TCWVEKMLTFHYINTGNSMAPVSVSEIWPNSFSPQTDKLFNILDVQTTGEC----- 858
QY      1021 EPPGPPVHPPELOHTNRLNGSNTQCOVR-----CHLG 1053
Db      859 -----HFENYQRYCALCQOKSAMQTLKGIYVFLSKTDKRLKYCIKADPHCINFLCNFG 911
QY      1054 QLAKGTEVSVGLRLIVHNEFFRRAKFKSLTVST--FEL--GTEEGS--VLQLTEASRW 1106
Db      912 KMEGSGEAS-----VHIQLEGPSILMDETSALKFEIRATGFPEPNPVRVLELNKENV 965
QY      1107 SESLLEVV--QTRPILISIMILIGSVLGGILLALVFCIMKLGFF--AHKKIPEEKRE 1162
Db      966 AHVLEGLHQRPKRYFTVIITSSLLGLIVLLISYVMWKAGFFKQYKSILOENNR 1025
QY      1163 E 1163
Db      1026 D 1026
```

Search completed: July 16, 2003, 07:56:45  
Job time : 67.9093 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

COM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:02 ; Search time 119.819 Seconds  
(without alignments)  
2006.842 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106

Sequence: 1 MELPFVTHLPLVFLTGLC.....GFAHKKIPBEKREKLEQ 1167

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rudent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacterioph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1853.5	30.4	1171	13	Q42094 gallus gall
2	1487.5	24.4	823	4	Q8WY18 homo sapien
3	1136	18.6	1161	11	Q9QY67 ratu
4	1097	18.0	1161	11	Q9WTV4 mus musculu
5	1089.5	17.8	1160	11	Q9R200 mus musculu
6	1069	17.5	1169	11	Q9QX44 mus musculu
7	1000.5	16.4	1151	11	Q9J130 ratu
8	997	16.3	780	13	Q06271 xenopus lae
9	984	16.1	1086	4	Q96H11 homo sapien
10	980.5	16.1	1196	13	Q98TF1 cyprinus ca
11	953	15.6	1167	11	Q88340 ratu
12	947.5	15.5	1187	13	Q98TF0 cyprinus ca
13	946	15.5	1167	11	Q88341 ratu
14	883.5	14.5	920	6	Q28984 sus scrofa
15	883	14.5	895	11	Q9WU68 mus sp. itg
16	797	13.1	1332	5	Q9BP08 halocynthia

17	778	12.7	1036	11	Q91YD5 mus musculu
18	732.5	12.0	1054	5	Q9UG61 strongyloce
19	725	11.9	1033	6	Q9BGJ3 bos taurus
20	691	11.3	1032	11	Q61989 mus musculu
21	681	11.2	1034	13	Q98RT7 gallus gall
22	661	10.8	1041	5	Q76378 lyechinus
23	655	10.7	1041	5	Q9UB90 lyechinus
24	623.5	10.2	1033	13	Q42598 xenopus lae
25	583	9.5	1016	13	Q91779 xenopus lae
26	576.5	9.4	974	11	Q924W2 ratu
27	570	9.3	1045	13	Q91292 pleurodeles
28	562.5	9.2	1049	4	Q96H45 homo sapien
29	560.5	9.2	1047	6	Q9MZD6 bos taurus
30	551.5	9.0	1007	6	Q9GK48 bos taurus
31	551	9.0	833	5	Q9BP07 halocynthia
32	536	8.8	1034	6	Q9TUN6 oryctolagus
33	524.5	8.6	1036	6	Q9TUN6 sus scrofa
34	516.5	8.5	1012	11	Q70304 mus musculu
35	515.5	8.4	1049	5	Q8SY51 drosophila
36	511	8.4	1087	13	Q91778 xenopus lae
37	469	7.7	1036	6	Q9TUN8 canis fami
38	461	7.5	1036	6	Q9TUN4 canis fami
39	450	7.4	1021	5	Q9GSF4 podocoryne
40	438	7.2	1086	5	Q18428 geodia cydo
41	394.5	6.5	191	6	Q29124 sus scrofa
42	374	6.1	764	6	Q28762 papio cyno
43	353	5.8	622	13	Q91570 xenopus lae
44	305	5.0	673	5	Q95YMO polyandroca
45	251	4.1	1472	13	Q90Z40 gallus gall

## ALIGNMENTS

## RESULT 1

ID	Q42094	PRELIMINARY;	PRT;	1171 AA.
AC	Q42094;			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	ALP1 integrin.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianine;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISUE=GIZZARD;			
RX	MEDLINE=97476270; PubMed=9334246;			
RA	Obata H., Hayashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,			
RA	Sobue K.;			
RT	"Smooth muscle cell phenotype-dependent transcriptional regulation of			
RT	the alpha integrin gene."			
RL	J. Biol. Chem. 272:26643-26651 (1997).			
DR	EMBL; AB000470; BAA3160.1; -			
DR	EMBL; AB000471; BAA3161.1; -			
DR	HSSP; R17301; IAOX.			
DR	InterPro; IPR00413; Integrin_alpha.			
DR	InterPro; IPR002035; VWF_A.			
DR	Pfam; PF01839; FG-GAP; 5-			
DR	Pfam; PF00357; Integrin_A; 1.			
DR	Pfam; PF00092; vwa; 1.			
DR	PRINTS; PR00453; VWFADOMAIN.			
DR	SMART; SM00191; Int_alpha; 4.			
DR	SMART; SM00327; VWA; 1.			
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.			
DR	PROSITE; PS50234; VWA; 1.			
KW	Integrin.			
SQ	SEQUENCE 1171 AA; 130228 MW; B5054AD65F09736E CRC64;			
Query Match	30.4%;	Score 1853.5;	DB 13;	Length 1171;

Best Local Similarity 34.7%; Pred. No. 2.5e-131;  
Matches 416; Conservative 226; Mismatches 464; Indels 83; Gaps 21;

```
13 LVFLTLCSPNIDENHPRLPFRPPAEFGYSLVGHVGGQGRMLVGAWMDPSGRRD 72
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
8 LVALPILIASVNDVKNAMTFSGLPDMFGYVQOYENEGKVLIGSLVPLVQPEKRTD 67
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
73 VYRCRPGAHNAPCAKHLGD-YOLGSSHPAVNMHIGMSLETDDGGFMACAPLMSRA 131
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
68 VYKCPGRDSQSPCITLNPDAISVNPWNEVENMILGTTIV-TNKSGFLACGPIYAK 126
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
132 CASSVSSGICARVDASFPQGSAPTAQRCPPTVMDVITVLDGNSIYPMSEVQFLRL 191
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
127 CRLHYTTGVCNSVSTFETVKAVAPVQECKTOLDIVILDGNSIYPMSEVTAFLNLS 186
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
192 VGLKFLDPEIOIVGVGVGESPVHEWSLGDFTKTEKVVAAAKNLRSREKPTQAIW 251
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
187 LKNMDIGPQOTGVIVQGVHEFYLTSTTEEMDAALRIKRGGTQMTALGIDT 246
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
252 ACTEGSOSHGSRPEAARLLVVTDSHDEELPALACAGRTYRGIAVIGHYLR 311
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
247 ABEATEAHGARQGVQKMWIVTDSHDNRLQVICKEDENIQRPAIALISYSG 306
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
312 QRPSSFLREIRTIASDPERFFNVTDALTDIVDALGDRIFGLESGHAENESSGLE 371
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
307 NSTEKFEVEIKSIASKPTEKEHFNVSDELAITVIALGERIFALEATTDOOASFME 366
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
372 MSQIGSTRKLDGILFGWVAVADMGSVLMLEGHRLPPMALEDEFPALONHAATL 431
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
367 MSQIGSAVYSDQWMLGAVGADNMGVVMVWDSDISPSNDFDRHSEKIEPLAAL 426
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
432 GYSVSMILRGGRRLFLSGAPRFRHKGVIAPOLKKDGAVVAOSLOGEOISYFSELC 491
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
427 GYTVASALPRGG-VLYIAGOPRNHGVIIYKM-EGREVKYLQRLKGGIOISYFGVIT 484
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
492 PLDTDRDGTTLVLLAAMPFLGPONKETERVYIYLVG-----QOSLTLQGT----- 538
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
485 TIDNRDSFTDLLVGAFTYMGTEKEEGKVYVVALNKTKEFQMSLEPIKOTCCSPLHG 544
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
539 -----LQEPPODARFEGAMCALPDLDNQGFADVAVAPALDEHQGLYLYHGTQSVRPH 594
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
545 DTCKVLKNEPCGARFETIAAAYKDLNDYNDIVIGSPLEDHGRGVVITYHGHNTISK 604
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
595 PAQRTAASMPHALSYFGRSVGDRLDDDDLDVDAVGAQAAILSSRPVHLTPSLEY 654
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
605 YQRIASGGDEKVKVFFGQSVHGENMDLNDGLIDVIGLGAALFMSDVAEVANSQGF 664
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
655 TPQAISSVQDRCRRRQGEAVCLTALCFQVTSRTPEGRMD-HQPYMFTASLDWTAGARA 713
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
665 MEKSNINIOQONCOINRKRTICINATICFK--TRLKSKEDIFFESSLOQWITLDAORQISRS 722
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
714 ARDGGGQRLSPRLRLSVGNVTCBOUHFHVLDTSIDLPRVALTYVFDALNTTKP--GPTL 771
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
723 LTFETHERMOK--NTTIKGSBCIKHNFMLDKPDQDSVKVLEF--NFSDPESGPVL 777
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
772 NEGSPSTLOKLVFSDKCPDNECVTDLVLYQVNMDIRGSRKAPFVVRGGRKRVLTSTLE 831
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
778 DSNLNSISIEYIFPTDCGAKKKCISDLALNVKASISAGDSSSPFYKSNNDPFTQLSK 837
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
832 NKKENAVNTSLIIFSRNHLASLP-GRSPIKVECAAPSAHARLCVGHVFPQTAKV 890
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
838 NKKDSAVNTRALVQVSPNIIIFAGIEDIQDSS-----CESNMNITCKGVFPFLPREBI 890
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
891 TLLLEPESCSLLSQVFGKLTASSDSLEENGTLOENTQTSAYIOYEHLLFSESESTH 950
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
891 SKRISFQFASVLLLENATVDVATSDSEBPETLRNRRGQVITPVYEEVGLLFSVFEKH 950
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
951 RYEVHPYGLTPVG-----PGPEFKTTLVONLIGCVVSGLLIISALLPAVAHGN---YF 1001
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
951 HYIIAANETIFPAINTEBOIGDEVILHYRIEKGHNHPMNLILQILYPRVTAKTYLLYI 1010
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1002 LSLGOVITNNAAC-----IVONLTLPQGPVHPBEILOHTNRLNGSNTO 1044
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

```
Db      1011 TLLSH--SQMAICKSSYEPVDHLKIGSGKSVYLPKIKEPTKDTI-----MECDTFS 1058
Qy      1045 COVVRCHLGQALAKTEVSVGLRLRVHNEFPRAKRSPLTVSTFELGTEGSLQLTFS 1104
Db      1059 CASINCALAP-SDISQVNSL--RWKPTTIKASHSLHLVVKALLRENSSLIRNDHQ 1115
Qy      1105 RMSSESLLEVQTRPILISLWILIGSVLGLLLALLVLCMLKGFPAHKKIPEEKREB 1163
Db      1116 KLETKIKISKEPPESVPLWVILPISFAGLIILALLIFALWKAGF---KRLPKKMEK 1171

RESULT 2
Q8WY18
ID Q8WY18 PRELIMINARY; PRT; 823 AA.
AC Q8WY18;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE MST018.
GN MST018.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=AORTA;
RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
RA Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,
RA Sun Y.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
RA Gao R.L., Qiang B.Q., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
RA Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF111799; AAL39001.1;
DR InterPro; IPR00413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 4.
DR PRINTS; PRO1185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 4.
SQ SEQUENCE 823 AA; 92672 MW; DE4E78079DCD4925 CRC64;

Query Match 24.4%; Score 1487.5; DB 4; Length 823;
Best Local Similarity 39.6%; Pred. No. 8.2e-104;
Matches 327; Conservative 151; Mismatches 308; Indels 39; Gaps 13;
```

```
372 MSQIGSTRKLDGILFGWVAVADMGSVLMLEGHRLPPMALEDEFPALONHAATL 431
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1 MSQIGFSHVVEDGVLGAVGADNMGNVLETSGAKYIPLRESLYKFPBELKNHGAYL 60
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
432 GYSVSMILRGGRRLFLSGAPRFRHKGVIAPOLKKDGAVVAOSLOGEOISYFSELC 491
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
61 GYTVASVSSROGRVYVAGAPRFNHTGVILFTMHNNRSLTIHQMRGQOIGSYFGSEIT 120
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
492 PLDTDRDGTTLVLLAAMPFLGPONKETERVYVVLVVGOSLLTLQGTLPQEP--QDARFG 550
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
121 SVDDIGDSDVTDVLLVGAAMYFN-BGEREGKAYVEL-RQNFVYINGILKDSHSYONARFG 178
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
551 FAMGALPDLDNQGFADVAVGAPLEDHQGALYLYHGTQSVGRPHPAQRTAASMPHALSY 610
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
179 SSIASVRLNDQSVNDVVGAPLEDNHAGALYIIRHFGFSGSLTKPKQITASELATGQY 238
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
611 FGRSVGRLDLDGDDLDVDAVGAQAAILSSRPVHLTPSLEYTPQAISSVQDRCRRG 670
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
239 FGCSIHGOLDINDEGLDILVAGALGNAYILMSRPVQVINAHLPEPSKINIFPHDCKRSG 298
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
671 QEAVLTALALCFQVTSRTPGRMDHQFYVRFTASLDWTAGARAAPFDSGQRLSPRLSL 730
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
299 RDACFLAFLCFPTPIFLAPHPQTTTGVIRVATYMDERRRYTPRAHLDDEGDFPTNPAVL 358
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
731 VGNVTCBOUHFHVLDTSIDYLRVALTYVFDALNTTKPQPVINEGSPSTLOKLVFSPKDCG 790
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
359 SGQELCEHINHVLDITAYVAVPVPFSVYSLEDDH-GPMLDDGKPTTLRLSVSPFWNCGN 417
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
791 PDNECVTDVLYQVNMNDI-----RGSRK-----APFVVRGGRKRVLTST 828
Qy      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Db 418 EDHECVDPDLVDARSDLPAMCYCQVLRKPAQDCSAVTLSPDTTFPIESTRQAVNEA 477  
Qy 829 TLENKRENAVNTSLSIIFSRNLHLASLTPORESPIKVECAPS--AHARLCVGHPIVQT 886  
Db 478 TLENNGENNAVSTVLNISQANQFASLIQKEDSDSIECVNERLQKVCNVSVPFIRA 537  
Qy 887 CAAKTFLEFEFSCSLSQVFGKLTASDSIERNGTLOENTAQTSANIQYEPHLLFSSE 946  
Db 538 KAKVAFRLDFFESKSIFFHLHLEIELAGSDSNERDSTKEDNVAIRFHLKYEADVLFTRS 597  
Qy 947 STLHRYEHPVGLP--VGPGEFPTTLRVONLGCYVSGLISALLPVAHNGVYFSL 1004  
Db 558 SLSHYEYKANSLEBYDYGPPFCIFRIQNLGLPFIHGMKMTTPIATRSGNRLKLT 657  
Qy 1005 SQVITN--NASC-IVQNTLTPPGPPVHPPELOHTNRNGSNTOCCOVVCHLGOLAKATEV 1061  
Db 658 RDLFLDEANTSCINIGNSTEYRPTVE--BDLRARQNLHNSNDVVSINONI-RIVPNOEI 715  
Qy 1062 SVGLRLVHNEFFRAKFKSLTVSTFELGTEGSLQLTASRMSSESLLEVYQTRPLII 1121  
Db 716 NFHLGNMLRSKLKLYKMKIMVMAALORQFHSPIFREEDPQRQIVFEISKQEDWQV 775  
Qy 1122 SLWILIGSVLGLLALLVFLCMLKGFPAHKKIIEEKEEKE 1166  
Db 776 PLWIVGSTLGLLALLVIALMKLGPPRSAR---RRREPGLD 816

## RESULT 3

090YE7 PRELIMINARY: PRT, 1161 AA.

AC 090YE7  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Alpha D integrin.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY;  
RA Gallatin W.M.;  
RT "Cloning of rat alpha D, a novel beta 2 integrin."  
RL Submitted (Aug-1997) to the EMBL/Genbank/DBD databases.  
DR EMBL; AF021334; AAF21241.1; --  
DR HSSP; P11215; IABX.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR Pfam; PF01839; FG-GAP; 5-  
DR Pfam; PF00357; Integrin\_A; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PRO1185; INTEGRINA.  
DR PRINTS; PRO0453; VWFADOMAIN.  
DR SMART; SM00327; VWA; 1.  
DR SMART; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
DR Integrin.  
KW SEQUENCE

Query Match 18.6%; Score 1136; DB 11; Length 1161;  
Best Local Similarity 30.1%; Pred. No. 7e-77;  
Matches 363; Conservative 208; Mismatches 486; Indels 148; Gaps 46;

Qy 20 CSPNLDEHRRPLFGPPAEAFGYSLQHVGGQRMVLGAPWDPSGDRGDVYRCFVG 79  
Db 17 CHGSLNDVEEPIV-F-REDAASFGQTVQF--GSSR-LVVGAFLVAVANVQTRLDLC--- 69  
Qy 80 CAHNAPCAKGHLGDYQLGNSHPAYNMILGMSLLETGDGGMACAPLMSRACGSSVPS 139  
Db 70 ---AP-ATGMCQPIVL-RSPLEAVNMSLGLSVTAIINNAQLLACGPTAQACVKNMYAK 123

Qy 140 GICARVDASFOPOGSLAFTACRPTV-MDVVIVLDGSNSI--YPMSEVOTFLRLVGLKF 196  
Db 124 GSCILLGSLQFIQAVAPASMEPCRGEMDIADLLDGSSSINQRPQAKDKDFVKALMBE-F 182  
Qy 197 IDEBOIQVLVOGESPPHMSLDFRTKEEVRAAKLSRREGRETTAQAIVACTEG 256  
Db 183 ASTSTL-FSLMQYSNILEKTHPTFTEFNKILDPQSLVDPVIOQL-TYTAAGIRTVMBEL 240  
Qy 257 FSGSHGRPEARLLVVVTDSEH-DGEBLPAALKACGAVTVYGIADVLRORPD 315  
Db 241 FHSKNGSKSAKKLLVTTDQKTRDPLEYSDVIRPADKAGITRALGVCAF---QEP 296  
Qy 316 SSFLREIRTIASDPERFFNVTDAAITDIALDGRIRFGLBSGSHAENESSFGLMSQI 375  
Db 297 TA-LKEMLNTISAPRQDHVFVGNFPAALRSIQROLEKIFALIEGQSASSSSFOHMSQE 355  
Qy 376 GFSTHRLKDGILFGMGAIVMGSSVLMEGHRLFP-----RMALDEFPALONHA 428  
Db 356 GFSSALTSDBGFLVAGVGSFSW-----SGAFLYPPVTRPTFTINSGEN-----VDMRD 403  
Qy 429 AYLGSVSMILRGGRRLFTSGAPFRHKGVIAPOLKKDGAVVAOSLOEOIGSYRGS 488  
Db 404 SYLGISTAVARWKVHSLIL-GAPRHQHTGKVILF-TQEARHMRPKSEVRGTIGSYFGA 461  
Qy 489 ELCPDLTDRTDGTTLVVAAPMFLGPONKETGRVYVYLV-GQOSLTLTLOGLQDEPPD- 546  
Db 462 SLCSVDVDRDSDTLVLIGAPHY--EDTRGGQVSVFPVPRVGRWQCEALTLHGQGHFW 519  
Qy 547 ARFGFAMGALPDNLQDGFADYAVAGAPLEDGHOGALYLHG-TQSGVRPHPAORIAAAMP 605  
Db 520 GRFGVALTLVDVNGDNLDADVAIGAPGEESRGAVYIFHGASRLIEIMPSFGQVTSQLS 579  
Qy 606 HALYFGSVGRDLDDDDLVYAVAGAGAAIILSRPIHILRPSLEVTQALSVQRD 665  
Db 580 LRLQTFGSGSLGGDLTODGLVDYAVAGQVLLRLSPULKVLSIRFAPMEVAKAVYQ 639  
Qy 666 CRRR-----GOBAVCTIALCFQVTSRTGRMDH-QFYRFASLDEMTAGAAAPDG 717  
Db 640 CWERTPTVLEAGENTVCLT-----YKSGPDLIGNVQGSRYVDLADDPGLISRAIFDE 693  
Qy 718 SGORLSPRRRLSVGNVNTCEQLHFRVLD-TSDYLRPVALVTVPAL-DNTYKP---GPVLA 772  
Db 694 TKNCCTLNCRKTLGLGD-HCEIVKLLLPDCEDAVSPILIRLNFSLVDSASPRNLHPVLA 752  
Qy 773 EGSPTSIQKVPFSKDCGPDMECYTDLVQVNMIRSGRKAPFVVRGGRKRVLSTLEN 832  
Db 753 VGSQDHLTASLPFEKCKQELLCEGDGIGSFNS-----GLQVLVVGSPPELTVTVYVWN 807  
Qy 833 RKENAVNTSLSIIFSRNLHLASLTPORES--PIKVEG-AAPSAHARL---CSVGHPIV 884  
Db 808 EGEDSYGLTVKFFYPAGLSTYRVTGTQOPHOYPLRLACEAPPAQOEDLRSSCSINHFIF 867  
Qy 885 QTKAVTFLEFEFSCSLSQVFGKLTASDSIERNGTLOENTA-QTSAYIOY----- 937  
Db 868 REGAKTTFTMIFDVSYKAFGLDRL--LLRAKASNNKPDNTKAPQLELVKXTVYTLI 925  
Qy 938 -----EPHLLFSG-----ESTLHRYEHPVGLPVGPGGEFTTLRLVONLGCYVSGL 985  
Db 926 SRODSTNHVNFSSHGGRRQEAHRYVNNLSPL-----KLAVRVNVWVPLVNLGV 977  
Qy 986 II-SALLPVAHNGNYFLSLSQVITNNAACIVQNTLTPPGPPVHPPELOHTNRNGSNTO 1044  
Db 978 AVMDVTLSSPQG-----VSCVSG--MKRPQNPDLTQIQRSVLDGSLAD 1021  
Qy 1045 QOVVRCHLGOLAKTEVSVGLRLVHNEFFRAKFKSLTVVSTELGTEGSLVQL--TE 1102  
Db 1022 CLHFRCDIPSDIDDELDFIRGNLSFGWASQTLQEKVLVSBALFTFDTSVYQLPQOE 1081  
Qy 1103 A---SRMSSESLLEVYQTRPLIISMLIGSVLGLLALLVFLCMLKGFPAHKKIIEE 1159  
Db 1082 AFLRAQVETTLIEYVVEPI---FLVAGSSVGLLALLVALLVLYKLGFP----- 1128

QY 1160 KRECK 1164  
DB 1129 KROYK 1133

RESULT 4  
Q9WTV4 PRELIMINARY; PRT; 1161 AA.

AC Q9WTV4;  
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
DE Integrin alpha L.  
GN ITGAL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxId=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=DA/2J; TISSUE=SPLEEN;  
RA Ma R.Z., Teuscher C.;  
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF065902; AAD25885.1; -  
DR HSSP; P20701; 1LFA.  
DR MGD; MGI:96606; Itgal.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01839; FG-GAP\_5.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha\_5.  
DR SMART; SM00327; VWA\_1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS00234; VWF\_A; 1.  
DR PROSITE; PS00234; VWF\_A; 1.  
SQ SEQUENCE 1161 AA; 128240 MW; 86B102F7B209E431 CRC64;

Query Match 18.0%; Score 1097; DB 11; Length 1161;  
Best Local Similarity 27.9%; Pred. No. 6.4e-74;  
Matches 358; Conservative 187; Mismatches 443; Indels 294; Gaps 49;

QY 12 PLVFLTGL-----CSPFNLDENHRLPFGPPEAFEGSYLVQHGQGMMLVGAPEWDGDS 66  
DB 8 PRLLILGLQLPKAWSYNDTRPTQSFQAQGRHFGYQLIDG-----VVVGARPE----- 60  
QY 67 GDRRGDVYRCVPGAGANAPCAKGLDYLQGNSSHPAVNMHGLMSLLETDGOGFMACAP 126  
DB 61 GNTGTGLYHCRTSSEFCQPV-----LHGSNH--TSKYLGMTLATDAKSLACDP 110  
QY 127 LMSRAGSSVFSFGIARVDASFQPGSLAFTACRCPT-----MDVVYLDGNSI- 178  
DB 111 GLSRITDQNTVYSGLC-----YLPQSLRGPMLONPAYOECMKGVLDVFLFDGSGQSD 165  
QY 179 -YPMSEVQFLRLVGLKFLIDPEQIOVGLVOYGESFVHMSLDFPTKEEVVRAAKNLER 237  
DB 166 RKDFEKILFEMKDVWKL--SNYSQFAVQFSTDRCTETFLDY-----VKQNNRPV 217  
QY 238 RRGRE-----TKTAQAINVACTEGFSQSHGRPREARLLVVTTDGSHDGEEPLPALA 291  
DB 218 LIGSVQPMFLNTPTFRAINVVAHVAFKESGARPDATKVLVITDGEADKGNISA---- 273  
QY 292 CEAGRTYRGIVLGVYLRORPSSFLREIRTAIDPRPFPFVNTDAALTDIYDALG 351  
DB 274 --AHDITRTIIGIKFVSVQK-----KTLHIFASEPVEEFKILDTTEKLDLFTDQ 326  
QY 352 DRIFGLEGSHAENESSFGLMSQIGFSTRHLKDGILFGWGAADMGSVYLMLEGHRLP 411  
DB 327 RRIYALEGINRQDLTSFNNELSSGSIADLSKGNAAVGAADKMGGLDLR----- 379  
QY 412 PRMALED-----EPFPAALONHAAYLGVSVMILRGGRLLFSGAPFRFRHGKVIAP 463

DB 380 -----EDLOGATFVQEBPLTSDVRGGLGYTVAMNTSSSRPLAAGAPRYOHVQVLLF 434  
QY 464 OL-KKQGAIVAVQSLQGEQIGSFGEELCPIDTDBDGTDDVLLVAAPMFLGPONKETGRV 522  
DB 435 QAPENAGRMNQOKIEGTQIGSYFGGELCSVDLDDGSAELLILAPLFFGEQ--RGGRV 492  
QY 523 VYVIVGQSSLTLOGLTQPEPPQD--ARFGFAMGALPDLDNODGFADVAVAPLEDGHQAL 581  
DB 493 FTY-ORQSLFEMVSELQDPCYPLGRFGAALITALTDINGDRLTVAVAPLFE--QGA 549  
QY 582 YIYHGTQGVRRHPQRITAASMPHALSYFGRSVUGRLDDGDDLDVDAVAGQAAILLS 641  
DB 550 YIFENKPGGLSPQPFQRIGQAVFPFGSHIGVADLGDDRADVAVGEGRVVLS 609  
QY 642 SRIYHLTPSLSEVTQQAISVQDRC-----RRRQGAVCLTALCFQYTSRTPGRDHOFY 697  
DB 610 SRPVVDVYTELFSFSEELPHEVEGYSAREQKIGVILKA--CRIRKPLP-----QFQ 662  
QY 698 MR-----FTASLDEMTAGARAAPDQSGQRLSPRRLSVGNV-----TEQLHFVH-L 744  
DB 663 GRLLANLSTYLLQDGHRRSRQLFPDGSHEL-----GNTSITPKKSCIDFHFHPI 714  
QY 745 DTSVLRPVALTVTALDNTTKPGVYLNESFT-----SIOKLVPEKSD 788  
DB 715 CIQDILSPINVSINPSL-----LEBEQTPROCKGRAMQILRPSIHTVTKELPFQKN 766  
QY 769 CGPDNECVTDVLYQNMIDIRGSRKAPFVVRGGRKVLVSTLLENKENAVYTSLSIFSR 848  
DB 767 CGEDKKCEANLTL-----SPARSGPLRLMS--ASLAVEMVLSNGEDAYWRDLDP 820  
QY 849 NIIHLASLTP-QRESPIKVECAAPSANARL-----CSVGHVPFQTAGAVTFLLEFEFSCS 901  
DB 821 GLSPFKVEMLOPHSHMPVSCCELTEGSSLLTKLTKCNVSSPFIKAGQVVS----- 870  
QY 902 SLLSQVFGKLTASS--DSLEBNGTL-----QENTAOQSAIYQVPHLFFSESSTL 949  
DB 871 -LQVMEFTLLNSSWEDVEVLNGTVHCENENSSLOEDNSAATHPLVLPVNLITKEQN- 927  
QY 950 HREYHVPYGLTPVGSGPEFKTLRLVQNLGCYVVSGLIISALLPAVAGNFFLSQYIT 1009  
DB 928 -----STLYISTPCKPQTOQVQVH-----YQVRQBPAY 957  
QY 1010 NNASCIQNLTEPPCP-----PV-----HPEELQHTNR-----LNGSNTQC 1045  
DB 958 DHNMPTLEALVGVPMHSEDPITYTWSVQTDPLVTCHEDDLKRPSSBAQCLPVGCR 1017  
QY 1046 QVVRCHLQOLAKGTEVSGLRLVHNEFFRAKFKSLTVSTFELGT--EGSVQLTREA 1103  
DB 1018 PIV-----FREIL--IQVTGIVELSKELIKASSTLSLSS 1050  
QY 1104 -----SRWSES-----LLEVQTRPILISLMLIGSVLGLLLALVLC 1144  
DB 1051 LSVFNSKSHRLVYSKASEAQLVKVDLHEKML-HVYVLSG--IGGLVLLFLIFLAL 1107  
QY 1145 WKLGFFAHKKIPEEKREKLE 1166  
DB 1108 YKVGFF-----KRLKEKME 1122

RESULT 5  
Q9R200 PRELIMINARY; PRT; 1160 AA.

AC Q9R200;  
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
DE Integrin alpha L.  
GN ITGAL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxId=10090;  
RN (1)

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=SPLEEN;  
 RA Ma R.Z., Teuscher C.;  
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF065901; AAD25884.1; -  
 DR HSSP; P20701; 1LFA.  
 DR MGD; MGI:96606; Itgal.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS50234; VWFA; 1.  
 SQ SEQUENCE 1160 AA; 128127 MW; A33C531B139F1EAD CRC64;

Query Match 17.8%; Score 1089.5; DB 11; Length 1160;

Best Local Similarity 28.3%; Pred. No. 2.4e-73;

Matches 359; Conservative 189; Mismatches 454; Indels 265; Gaps 49;

12 PLVFLTGL-----CSPFLIDEHHPRLFPGRPEAEFGSYLQHVGGGQRMVLGAPWDGPS 66  
 8 PRLLLGLQLFAKASYNLDTRPTQSLAQGRHFQVQLQEDG---VVVGARGE--- 60  
 67 GDRRDVRCFPGVGAHNAACAKGHIGDYQNGSHPAVMHLMGLLLETDDGGMACAP 126  
 61 GDNLTGLVHCRTSSEFCQFVS-----LHGSMN--TSKYLGMTLADDAKSLAADP 110  
 127 LMSRACGSSVFSSGICARVDASFOFGSLAPTORCPTY-----MDVIVLIDGSNSI- 178  
 111 GLSRICDONTYISGLC-----YLPQSLBGMLOKRPAYQECMKKQVLDVLPDSQSLD 165  
 179 -YPMSEVOTFLRLVGLFLIDEQIQVGLVQVGGSPVHEWSIGDFRTKEEVRRAKNSLR 237  
 166 RKDFEKLTFMDVNRKL--SNTSYQFAVGFSTDCRTEFTPLDY-----VKQKNBDV 217  
 238 REGRE-----TKTQALIMVACTEGFSQSHGGRPREARLLVVTGESHDELPALAKA 291  
 218 LLSGVQPMFLTLNTRAINVVAHAVFKESGARPAKTKVLTITDGEASDKGNISA--- 273  
 232 CEAGVTVYGIYAVLHRYRORDPSSFLREIRTIASDPDERFFNVTEAALTDVDAIG 351  
 274 --AHDTIRYIIGIGHFVSVQKQ---KTLHIFASEPVEEYVKLIDTFEKLKDLFTDQ 326  
 352 DRIFGLSGSHAHENESSFGLEMSQIGFSTRRLKDGILFGVNGAYDWGGSVLMLEGHRLFP 411  
 327 RRIYVIEGTNRQDLTFNMNELSSSGISADLSKGHNVGAVGAKDWAGGFLDLR----- 379  
 412 PRMALED-----EPFALONHAAYLYGYSVSMILRGGRPLFSGAPRFHRSKVIAF 463  
 380 -----EDIGATFVGOEPLTSDVRGGYLYTVAMWTSRSRPLLAAGAPRYHVGQVLLF 434  
 464 QL-KKDGAVRVAQSLQGEQIGSYFSGELCPLTDPTDVTLLVAAPFLGQNKETGRV 522  
 435 QAPBAGKRNQTKIEGTQIGSYFGEELQSVLDQGEHELILGAPLFPFGQ--RGGRV 492  
 493 FTY-QRQGLFEMVSELQDGPYIGRFAAITALTDINGDRLLTVAVGAPLEE--QGANV 549  
 523 VYVIVGQOSLTLTQGLQPEPPQD-ARFGFAMGALPDNLQDGFADVAVGAPLEDHQAL 581  
 582 YLYHGTQSGVRHPQRILAAASMPHALSTFGSGVDCRLDDDDLVDAVAGQAAIILIS 641  
 550 YIFNKRPGGLSPQRIQGAQVFPDIRWFGRSIGVXKDLGDBRIANVAVGEGGVVVIS 609  
 642 SRPIVLTSLFVTPAIVSVQDQ---RRGOAEVCTTALCFQVTSRTGRMDHORY 697  
 610 SRPVDVYVELSFSPEIIVHVEGYSKAREQKGVKAKA--CRIRKILTP-----QFO 662  
 698 MR-----FTASLDWTAGARAAFDGSGGRLSPRLRLSVGNV-----TCBQLAHFV-L 744

DB 663 GRLLANISYTLQDLGHNRMRSGLFPDGSHELS-----GNTSITPDKSLDFHFPPI 714  
 QY 745 DTSYLRVALVTVFALDNTTKPGVLINEGSPT-----STOKLVFESKD 788  
 DB 715 CIQDLISINVSINFSL-----LEEETPRQOKGRAMOPIRLPSHYTKELIPEKN 766  
 QY 789 GCPNCECVTLVQVNMDIRSGRAPVVRGRRKVLSTLLENKENAYNTSIIISR 848  
 DB 767 CGEDKKCEANLTLS-----SPARGPLRIMS-ASLAVEMWTLNSGEGDVAVVRDLDPFR 820  
 QY 849 NLHLASLTP-QRESPIKVECAAPSAHARL-----CSYGHVFPOTGAKVTLLEFEBSGS 901  
 DB 821 GLSFRKVEMLQPHSRMPSPCELTGESSLTKTLKCNVSSPIFRAQGEVS----- 870  
 QY 902 SLISQVREKLTASS--PSLENGTL-----QENTACTSAVIOYEPHLFSSSESTL 949  
 DB 871 --LOVMEFTLINSWEDVEVLGVHCENENSSLOEDNSAATHLPVLYPVNILTKEDN- 927  
 QY 950 HRYEVHPYGLTPVGPGDEFKTLRVQNLGCYVSGLIISALLPVAHAGNYFLSLQVIT 1009  
 DB 928 -----STLYISFTTPRGKPTQOVQHV-----YQVRIQPEAY 957  
 QY 1010 NNASCIVQNLTPRPGPPVHPEEL-----QHTNRLNSNTQOVVRCGLQGLAKTEVS 1063  
 DB 958 DHNPMTLEALGVDRP--HSEDLITYTWSVQTDPL-----VTCHSEDLKPPSSSEAE 1006  
 QY 1064 GLRLVNHNEFFRRKFKF---SLTVSTFELCTE--EGSVLOLTERA----- 1103  
 DB 1007 PCLPGV-QFRCPVFRWEIILIQVTVELSKEIKASTLSLGSLSVSNSSKHFLYG 1064  
 QY 1104 SRWSES-----LLEVQTRPILISLMILIGSVLGGILLALVFCMLKGFPAHKIPEE 1159  
 DB 1065 SKASEAQVLVAVDILHEKEML-HYVVLSG--IGGLVILFLIFLALYKGF-----KR 1114  
 QY 1160 KREEKLE 1166  
 DB 1115 NLKEKME 1121

RESULT 6  
 Q9OXH4 PRELIMINARY; PRT; 1169 AA.  
 AC Q9OXH4;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, last annotation update)  
 DE Leukocyte adhesion glycoprotein p150.95 alpha integrin subunit.  
 GN ITGAX.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Huang X., Gorski K., Tong C., Rattis F.-M., Teeng S.-Y., Pardoll D.,  
 RA Tsuchiya H.;  
 RT "Isolation of Genes Selectively Expressed by Dendritic Cells";  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF211864; AAF23492.1; -  
 DR HSSP; P1215; 1ABX.  
 DR MGD; MGI:96609; Itgax.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 4.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; UNKNOWN\_1.  
 DR PROSITE; PS50234; VWFA; 1.

```
KW Integrin. 1169 AA; 129150 MW; C616412033C219A6 CRC64;
SQ SEQUENCE

Query Match 17.5%; Score 1069; DB 11; Length 1169;
Best Local Similarity 28.3%; Pred. No. 8.6e-72;
Matches 354; Conservative 213; Mismatches 490; Indels 196; Gaps 49;

QY 1 MELPETHFLPLVFLTGLCSPPNLDENHRLPPGPEAEFGYVLQHVGGGQRMWLVGA 60
DB 1 MCGTWAIFLL-LGFVS--CLGPNLDAEKLTHP-HNDGAEFGHSVLQY---DSSWVVA 53
QY 61 PWDGSGDRRDYRCRPGVGAHNAAPCAKGLGDYQLGNSSHPANNHLMGSLLETDGCG 120
DB 54 PREIKATNOIGGLYKC---GYHTGNCPEPISL-----QVPEAVNISLSTLSLAATNP 104
QY 121 FMACAPLMRACSSVSFGSGICARVADSFQPGSLAPTQR--CPHY-MDVVIVLDGNSI 178
DB 105 LLAAGFTVHTHCENIYLTGLCLFLSSFR-QSQNPFAQOECPKODDITVFLIDSGSI 163
QY 179 --YPMSEVOTFLRLVGLKFLIDPEQIOVGLVOYGESPVHWSIGDPRTEKEVVARAKNLS 236
DB 164 SSTDFEKMIDFVKAWSQL--QRPSTRFLMQSDYFRHFTFNNTISTSSPLSLGSR 221
QY 237 REEGRETQAQAIMVACTEFGFSQSGRPEARLLVVTDSGH-DGEELPAALKACEAG 295
DB 222 QLRG-VTYTASAIKHVITELFTTQSGARODATKVLIVITDGRKQDNLSYDVIWAEAA 280
QY 296 RYTRVGIAVLGHYLRQRDPPSSFLREIRITASDPDRFFNFDEALDIDVALGDRIF 355
DB 281 STIRTAIG-GKAFVNEHSK---QBLKAIASMPSEHYFVSENPALDIEQLEKILF 335
QY 356 GLEGSHAENESSFGLEMSOIGFSTHRLKDGLFGVGAVDWGGSVLMLEGHRLPEPMA 415
DB 336 AIBGTTPSSSTFELMSQEGFSAVFTPPGVLAAGGSSW-----SGCAFVLYSNMR 388
QY 416 LDEFPALQNH-----AYLGYSVSMILRGGRRLFLSGAPFRFRKVIAPQLK 467
DB 389 -----PTFINMSQENEDMRDALVGYSTALAEFKVHSLIL-GAPRHQHGKVIF-TOE 440
QY 466 DGAVRVAOGLQGEQIGSYFSGELCPYDTRDGTDTVLVAAPFLGPKETGRVYVY-L 526
DB 441 SHHMRKSEVRGQIQSYFGASICSVDMDRDSITDVLGVPHYV--ETRRGQSVSCM 498
QY 527 VGOOSLLTQGLQPEPPQD--ARFGAMGALPDLNDGFADVAVGAPLEDGQALYLH 585
DB 499 PGVSGMHCQGTLLHGGQHPWGRFGAALVTLGVNDSLADVAIGAPGEENRGAVYIH 558
QY 586 G-TQSGVRPHAPORIAAASMPHALSTFGASVDRGLDDGDDLVDAVAGQAAIILSRP 644
DB 559 GASRODIAFPSORISASQIPSRIOYFGQSLSGGQDLTRDGLVDLAVGSKGVLLLRTP 618
QY 645 IYHLPSLEVTPOAISVQDRCRR--GOEAVCLTALACQV--TSRRPGMWDQFVYRFTA 702
DB 619 IIRVSTVFHTPREISRSRVCEQOVARBOTLSDATVCHIHESPKTQJGDRSTVFDL 678
QY 703 SLDEWTAGARAAFDGSGQRLSPRLRLSYGNVTCEQDHFHV-LDTSYLRLPALVTFTAL 761
DB 679 ALDHGSLRRAIFKEXKTRALTRVKTLGL-NKGCESVKLLIACVDSVTPILRLNFSL 737
QY 762 ----DNTTKPGVNLNESPSTSIQKLVPSKQCGPNECVTDLVLQVNDIRGSKAPV 816
DB 738 VGVPISSLQNLQMLAVDDQYFTASLPFEKNCGADHICODDLV-----VEGFPLKLT 792
QY 817 VEGGRKXVIVSTLLEKRENAVNTSLTIF-----SR 848
DB 793 VVGSDELAVNDVVSNDGEDSYGTYTLVYPVGLSFRVAEGVLRKKEQDQOMORRGH 852
QY 849 NLHL-ASLTPORESPIKVECAAPSAHARLCSVGHPIYOTGAKTFLLEFSCSILSYQ 907
DB 853 SLHLMODSTFPRDSQGL-----WSTSCSSNHVIFRGGSQMTFLVTVDVSPKMLGR 903
QY 908 FGLTRASSLSLEBN--GLQENTAOQTSAYIOYEPHILFSESTL----- 949

DB 904 L--LILRAVGSBNVPGT-PKTTFQLELPVKAAYVYTMISSHDQTKYLNFTSEKEXTSV 960
QY 950 --HRYEVHRYGTLPGPGPEFKTTLRQNLGCVYVSGLIISALLPAVAHGVNPLSLSOV 1007
DB 961 VEHRFQVNNLGRDVPVAINFWVPELKEGAVWVY-----MWHHPON---PLQOC 1007
QY 1008 ITNNASCIQVNLTEPPGPPVHPEELOHTNR--LNGSNTQCVVACHLQOLAKGTEVSG 1064
DB 1008 YRNRLK-----PTQFDLLTHMQKSPVLDGS IADCLHLRCDI-----PSLG 1047
QY 1065 LIRLVHNEFFRAKKSITLVSTFELGTEBGSVQLTA-----SRMS-----ESLLEV 1113
DB 1048 ILDELY--FILKGLNLSFGWISQTL-----QKQVLLISEALETNTSVYSQLPGQDAFLR- 1099
QY 1114 VQTRPIL-----ISLWILIGSVLGLLILALLVFCMLKLGFF--AHKIPKE 1158
DB 1100 AOTKIVLEMYKHNVPPLVGVSSVGGLLLILITAILYKAGFFKQYKEMLEE 1152

RESULT 7
09J130 PRELIMINARY; PRT; 1151 AA.
AC 09J130;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Integrin beta 2 alpha subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Fathallah D.M. Sr., Zeria K. Jr.;
RT "Cloning of the rat CD11b cDNA sequence.";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF268593; AAF81280.1; -.
DR HSSP; P11215; IABX.
DR InterPro; IPR00413; Integrin_alpha.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF01839; FG-GAP; 5_A.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA_1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 1151 AA; 126943 MW; 8F785695D4074C45 CRC64;

Query Match 16.4%; Score 1000.5; DB 11; Length 1151;
Best Local Similarity 28.4%; Pred. No. 1.3e-66;
Matches 346; Conservative 216; Mismatches 491; Indels 165; Gaps 45;

QY 19 LCPSPNLDENHRLPPGPEAEFGYVLQHVGGGQRMWLVGAPMDGSGDRRDYRCPV 78
DB 13 LCHGNLDTENPMFTQ--ENAGSFGQSYQL---GETRVVAAPQEVKAVNGTALQYC-- 66
QY 79 GGAHNAAPCAKHLGDYQLGNSSHP-----AVNMHLMGSLLETDGCGFMACAPLMSR 130
DB 67 -----DYST-NRCDPRLQVPEAVNMSLGLSLAATTVPPOLLACGFTVHQ. 111
QY 131 AGGSVFSFGSICARVADSFQPGSLAPTQR--CPHY-MDVVIVLDGNSI--YPMSEVOT 186
DB 112 NKEKVTYVNGCYLVFGSVLRLKPOQFPALGCPQOESNIAPFLIDGSGSINTTIDFQKKE 171
QY 187 FLRLVGLKFLIDPEQIOVGLVOYGESPVHWSIGDPRTEKEVVARAKNLSREGRRTKA 246
DB 112 FVSTYMDQ--FQKSTLFSLMQYSDPEFTHHTFNDFKNPKPKSHVRIQLNGR-TKTA 228
QY 247 QAIMVACTEFGFSQSGRPEARLLVVTDSGHDEEL---PAALKACEAGRTVRYIA 303
```

DB 229 SGIRKVRBELFOKINGARDNAKILVITDGEKF-GDPLNTEDVLPBEAEAG-TIRYIIG 286  
QY 304 VLGHVLRORDDSSFLREIRITIASPDERFFNNVTDEALTDIVDALGDRIFEGSGSHAE 363  
DB 287 V-GNAFHK---POS-RRELDITASKRAGDHVQVONFELNLTNRQLOEKIATIGSTQC 341  
QY 364 NESSFGELESGQIGFSTRHLKDGILGSMGAYDM-GGSVLMLEGHRLPPRMALDEPPP 422  
DB 342 STSSEHEHMSQEGFSAITNSGPLGSGVSPDMAGAFLLPSKDXASFTNTTRIDSDM- 399  
QY 423 ALQNHAAVGYSSVSMLLSGRRFLSGAPRRHKGKVAIFOLKXDGAVRVAOSIQEOI 482  
DB 400 ---NDATLGTG-SAVISRNVOVLGAPRYOHGLVWFK-QNFGAHEPHTDIKSGCI 453  
QY 483 GSYFSGELCPDTRDGTDLVLAAPMELGQNKETGRVYVYLVGQ---QSLTLTQ 537  
DB 454 GSYFASLCSYMDADGNTNLLIGAPHNY--EKRRGQVASCPLPRGARQCCAILHG 511  
QY 538 TLQPEPPDARFGFAMGALPDINOGFADAVAGARLEDHOGALYVHGTQ--SGVRPH 595  
DB 512 D-QGHP--WGRFGALTVLGDVNGDKLIDVAIGAGEOEONGAVYIFHGASVASISTPH- 567  
QY 596 AORIAASMPHALSYGRSVDRDLDDDLVDVAVAGQAAILLSRPIVHLTSLMT 655  
DB 568 SORIGARPSPGLQYFGOSLSGKDLTMDGLMDLVAGQGRLLLRADQVRLKEXMPS 627  
QY 656 FOAISVORDCRRR-----GQEAVCUTALCFQVTSRTPGR--WDHQFYMRFTASID 705  
DB 628 PGNLARSPVACQEEVYVKNKDAEVAVCL-----QVRKTKRLKEGDIQSVVYTDLALD 681  
QY 706 EWTAGARAAFDGSGGRLSPRLRLSVGNVTCQOLHFHVLDTSD-VLRPVALTPTA--- 760  
DB 682 PGRSVVRAFFEDTKNGIL-RRIRVFLGLOKCELTKLILPDQVDNVSPIITRLNTLVGE 740  
QY 761 -LDNTTKGQPVLANESPTSIOKLVPESKDCGPDNCTVDLVQVNMDIRGSKAPFVVG 819  
DB 741 PRRSRDRLRPVLAEMAQRIFTAMPPEKXGMDTICODDLSTIVS---STRVNTLVVG 795  
QY 820 GRRKVLVSTTLERKRNAYNTLSIIFSRHLH---ASLTPQRESPIKECAARSA--- 872  
DB 796 DSRDDVAVTLRNDDEDSYGTCTCYRBSGLSYRKVASQONOSKPKMRV-IAEFSSEB 854  
QY 873 ---HARLCSVGHVFOGAKVTPLEEF-SCSSLLSQVFGKLTASSDSIERNTLOEN 927  
DB 855 QGVLTSTIWDINHPIFPANSEVTFVNTFVDSDASLKNKLLKVVTS---ENNVSRDIX 911  
QY 928 T-AQTSATIQYEPHLLFS-ESTL-----HRYEHPHG-TLPYGPPE 967  
DB 912 TESQLELPVKATVAVVTSSESSIKYLNFTASEMTSKYIOHOYQFNMLGORSFPVS--- 967  
QY 968 FKTLRVONLGCYVSSGLIISALLPAVAHGNYFSLSOVITNNASCIVONLTEPPGPV 1027  
DB 968 -----VFWIWOINKVTIMPPPOVTSQONLSSVCRTQKSPSISK 1008  
QY 1028 HPEELQHTNRNLNGSNTQOVNACHLGOLAKGTESVGLRLVHNEFFRRAKKSJLVST 1087  
DB 1009 FODELERPVNLCSAVACKRIQCDIPFSKKEIFNVTLQGNLLPFWYIETSHDLVST 1068  
QY 1088 FELGTBESVLOLTASKRSESLLEV-VOTRPIILSLMILIGSVLGLLLALVFCIMK 1146  
DB 1069 AEILFNDAFALLPQOETFEVKAQETETKVEPYVHNVPPLIVGSSVGLVALLITAGLYK 1128  
QY 1147 LGFFAKKIPBEKREK 1164  
DB 1129 LGFF-----KQYK 1137

RESULT 8  
ID 006271 PRELIMINARY: PRT: 780 AA.  
AC 006271: 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE Integrin alpha-2 subunit (Fragment).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
NCBI\_TaxID=8355;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Meng F., Desimone D.W.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 95-168 FROM N.A.  
RX MEDLINE=94008528; PubMed=8404528;  
RA Whitaker C.A., Desimone D.W.;  
RT "Integrin alpha subunit mRNAs are differentially expressed in early  
RT Xenopus embryos.";  
RL Development 117:1239-1249(1993).  
DR EMBL; L43058; AAA69770.1; -;  
DR EMBL; L43058; AAA69770.1; -;  
DR HSSP; P11215; IABX.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR Pfam; PF01839; FG-GAP; 4.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR PRINTS; PRO1185; INTEGRINA.  
DR SMART; SM00191; Int\_alpha; 4.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
FT NON TER 1  
SQ SEQUENCE 780 AA; 87017 MW; 9518B18CB2B6PF637 CRC64;  
Query Match 16.3%; Score 997; DB 13; Length 780;  
Beet Local Similarity 31.6%; Pred. No. 1.3e-66;  
Matches 253; Conservative 161; Mismatches 338; Indels 48; Gaps 17;  
QY 391 VGADWGSVYVMEGGRHLPPEPMALDEPPALQNHAAVYGYVSSMLRGGRFLPSG 450  
DB 1 VGADWGSVYVMEGGRHLPPEPMALDEPPALQNHAAVYGYVSSMLRGGRFLPSG 450  
QY 451 APRRHGKVAIFOLKXDGAVRVAOSIQEOIGSYFSGELCPDTRDGTDLVLAAPM 510  
DB 58 APRIDYGVQVYVYVITSGANVSVIOQREGIOISYFGSVLCSVYVNRDSTIDVLLVGAFT 117  
QY 511 FLGPQNKETGRVYVYLVGQOSLLTLQGTLOPEPPDARFGFAMGALPDLDNODGADVAVG 570  
DB 118 FMNEYKKEEGVYVWFSTRDGLVQREGLEGPKSLENTFRFGSAIYELSDIDGYNDAVYG 177  
QY 571 APLEDHOGALYVHGTQSGVRPHPAQRIAAASMPHALSYGRSVDRDLDDGLDYLVA 630  
DB 178 APLENONGSAIYINGNKNTIRTYYSOKILGSSFNPGLOYFGRSVDGHODINGDTINDVS 237  
QY 631 VGAOGAAILLSRPIVHLTPELEVTPOAISVORDCRRRQGEANCLTALCFQVTSRTPG 690  
DB 238 VGAGKVIQILMSQVADVAVKALFTPOQIVLTNKA- -----VTRICFSAVFR-PA 288  
QY 691 RMDHQFYMRFTASIDETAGARAAFDGSGGRLSPRLRLSV---GNVTCQOLHFHVLDTSD 747  
DB 289 NSQSVAVTYATLADADLLSRVSRQGFRESNRRFLORNIVIGPTSCFEHVFVQETS 348  
QY 748 DYLRPVALTPTFALDNTTKG--DVINEGSPTSIOKLVPESKDCGPDNCTVDLVQVNM 805  
DB 349 DSENALRLINVAQN---PSSSPVLNYPSSSTEWEPFIPKDCGDCKICISDLSLOI-Q 404  
QY 806 DIRSGRAPFVYRGRKRVVSTTLERKRNAYNTLSIIFSRHLASLTLPQRESPIKY 865  
DB 405 QIPSDIKSPYIVSMKDRRLFOVLTNKLKVAAYTRNALFSENLFPASSTIPPDG-TEV 463  
QY 866 ECAAPSAHARL-CSVGHVFOGAKVTPLEEFSCSSLSOVFGKLTASSDSIERNTL 924  
DB 464 LCQVGNQNSVSCILIGFPLKDSOKVFDLWFDNANLNTAIFIQATISES--REARE 521  
QY 925 QENTAQTSATIQYEPHLLFSSESTLHRYEVHPGTLPV-----GCPPEFKTLRVONLG 978  
DB 522 EDNSVNTLTPVOYDAEHTLRTVNTINFEYEVFPGINIPSVVNTPDIDGLEFNFTKV-STG 580







OC Cyprinidae; Cyprinus.  
 OK NCBI\_TaxId=7962;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PERITONEAL EXUDATE CELLS;  
 RA Kimura M., Nakao M., Miura C., Fujiki K., Yano T.,  
 RT "Molecular cloning of a leukocyte integrin from the common carp."  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB048536; BAB39134.1;  
 DR HSSP; P20701; 1LFA.  
 DR InterPro; IPR001969; Asparticase site.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 SQ SEQUENCE 1196 AA; 132477 MW; 9369C807E7DCA53B CRC64;

Query Match 16.1%; Score 980.5; DB 13; Length 1196;  
 Best Local Similarity 27.1%; Pred. No. 4.7e-65;  
 Matches 334; Conservative 213; Mismatches 526; Indels 159; Gaps 48;

5 FVTH-----LFLPLVLTGLCSFNLDEHPRLPFRPPRAEFGSYLQHVGGGRMM 56  
 6 FVTQTSSRMETLGLFLFWMAISLSEAFNIDTEHPLRFNGAPEDFGSYVQTEFGNRCOI 65  
 57 LVGAWPDGSPGDRDYYRCVPVGAHNAFCAGKHLGDIYQGNSSHPAVNMHLSLETD 116  
 66 IVGALLEGNS---TEEMVST---ADLQSCQR-----LQPGSESVRF-FGMAAAS- 110  
 117 GDGGMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQR-PTMYDVIVLDGS 175  
 111 -SALTSCSPYPHEDCDNSYLVNGVCYFSSLSLQAVSNFTAYQCSKREVLVLFDS 169  
 116 NSI--YPMSEVQTFILRLVGLKFLIPDEQVGLVQYGESEPHENSLGPRT---KEVVR 230  
 170 SSMKTVDDEMNKNFTKIDIMKKL--SNSSTIKFAVQFSTDVTFVFDNDVQSSAEEKMK 227  
 231 AA-KNLRRREGREKTAQAIWVACTEGF-SQSHGGRPREARLLVVTVDGESHDEBLPA 287  
 228 ETHMKSLSL-----INTKALDIYILKNLNSMLSGADSKAQALVITDGDSDMDYD-N 279  
 288 ALKACEAGRVTRYGIAGVILGHVLRQRDPSSFLREIRTIASDPDERFFPNVTDEALTDIV 347  
 280 VLKCCDEQNIIRYIIGV-----GKVDLIELTQLASEPKRNTFYIKYSGKGLL 329  
 348 DALGRTIRGLSESHENESSFGLMSQIGFSTHRLKDGILFPMVGAAYMGSSVLMLEGH 407  
 330 DNLQKKINIEGSSVDAQGRDROKELSSQSFVVOEESYIVGVSNSMDRGALYVTSQ 389  
 408 RLFPRLMLDEFPALONHAAYLAGSVSMILRGGRLLFLSGAPRFRHGVIAFOLKK 467  
 330 SDFK-----ETIIPAVAKDSYMGYS-TVLKMRGVSILSGAPRAEHTGLVTLF-TKN 442  
 468 DGAVVAOSLOGEQIGSYFGESELCPLDDRDGTTDVLVAAPMFLGPONKETGRVYVVLV 527  
 443 QNTVTVTSNINGEQIGSYFGASLILDDVSDSDSDFLVGAPLFIQSGPRTEGRLVYVSL 502  
 528 GQGSLLITLQGLTQPEPPDARFGFAMGALPDLNOCGFADVAVGALPELDCGALVLYGT 587  
 503 SEQKY--FQKTLNVSQSTTGRPAASVASLSKLDINGGLSVAVGAPLE--NEGAVVYIYIGD 558  
 588 OS-GVRPH-PAORIAASMPHALSYFGREVDRLDLDGDLVDVAVGAQAGAILLSRPI 645  
 559 RHGHNELPTEPRISVQSVLPGLQPGVSLTQGMNNDNLTDIYIGAOGGIVLLKARV 618  
 646 VHLTSLLEVTPQAIIVVQDRCCRROGEAV-CLTALCFQVTSRT--PGRMDHQFYMRFTA 702

DB 619 MSVSAQLSPFSPEKISLNFEC--PGSNAFNAFNLTSCFTVTERSTSLKKUNVSNL 676  
 703 SLDEWTACARAFPGSGGRLSPRRLRLSV---GNVTCQLPHFVLD-TSDYLRPALVTV 758  
 677 NVDVYRGMSRGFPDQS--VSSRTIQSVLLDSSGSCNFSIFMLRCVADVSPKLKMN 734  
 759 PALDNTTKPG--PVINEGSPYSIOKLVFSSXDCGPDNCEVTDVLVQ--VNMDIRGSRKA 813  
 735 FSQTEMLSGNVAVLDVHSRTEENVEVFPQNCNNSNSVADLKLNFSTNNTLVENQA 794  
 814 PFVYRGGRKVLVSTLENKRENAVNTSLIIFSNHLASTLPQRESDPIVECA---AP 870  
 795 HFTVQ-----VSLNPDGDSYNTSIVLHYEGSLSKFDKIKPRTSSCCGRDSC 845  
 871 SAHARLCSGHPVEQTAQVTFLEFEF---SCSSLSGVFGKLTASDSLERGTQE 926  
 846 ATRRTGSDIDLPVRSIGTTTQFLGTFVYMKNDNSNMEIM--ITANSN--NGNRSD 900  
 927 NTAQTSAYIQYEPHILFS---SESTLHRYEVHPYGLPVGPGEFTKTLRVQNLGYV 982  
 901 TEVRSVAVQAVDLAISVAEDSVTNNFLEDRGPKPL-----NTYKVENSGRDL 954  
 983 SGLIISALLPA-----VAGGNVFLISQVITNNACIVON--LTBPG 1024  
 955 P-VSVTLTLPOTPHVILTPHTFSWHEVHH--SFISYHQI---MCLINKHLFFSP- 1006  
 1025 PPHVPEELQHNRLNGSTQGVVRCHLGQAKREVSGL--LRVH-NF----- 1073  
 1007 -----ELSAVOOVRRSTGWSIREVCSQFDLKKSAVFNLTADARLONVKEYSFY 1060  
 1074 -FRAPKSLTVSTFELGTEGSLQTEARSSESILEVQTRPILISIMILGSVIG 1132  
 1061 EFRKDNVSIASALNYSNLNQTSSBLKYNPHSQTEVKEFVPPSLMIVCGAV-G 1119  
 1133 GLILIALIVFLMKLGFFAHKKIPDEEKREK 1164  
 1120 GFFFLIILFLFLKCGFFKRRR-PDEFVQEDE 1150

RESULT 11  
 088340 PRELIMINARY; PRT; 1167 AA.  
 AC 088340;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Integrin alpha E1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98056820; PubMed=9394838;  
 RA Brenan M., Rees D.U.G.;  
 RT "Sequence analysis of rat integrin alpha E1 and alpha E2 subunits:  
 RT tissue expression reveals phenotypic similarities between  
 RT intraepithelial lymphocytes and dendritic cells in lymph.",  
 RL Eur. J. Immunol. 27:3070-3079(1997).  
 DR EMBL; AF020045; AAC23662.1;  
 DR HSSP; P11215; IABX.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 4.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 4.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS50234; VWF\_A; 1.

SQ SEQUENCE 1167 AA; 128970 MM; D88A2C38ACDC2AA8 CRC64;  
 Query Match 15.6%; Score 953; DB 11; Length 1167;  
 Best Local Similarity 28.1%; Pred. No. 5.5e-63;  
 Matches 367; Conservative 195; Mismatches 442; Indels 302; Gaps 65;

QY 14 VPLTGLC-----SPFNLDENH---PRLPPEPEAEFGYSVLQHV--GGGGRMMLVGP 61  
 DB 4 LPTFTLLCMASLKRQAGFNMDDWAWTALQPGAPAV---LSSLHLHDPSSNNQCLIVAR- 59  
 QY 62 WGPCSDRGDVRGCVG--GAHNAPCAK--GHL---GDYQLGNSHPAVNMHLGMSLLET 115  
 DB 60 --RSSNRNTALYRCASISIPDEIACQVPHICMPKGRYO-----GVLTV-- 102  
 QY 116 DGGGFMACAPLMSR-----AC-----MDVIVLDGNSIYPMSEVO----- 140  
 DB 103 GHNHGVLVITQVQARFRSLNSELTGACSLTTPNLDLQAOAFPSDLEGLDPCAHVDSGD 162  
 QY 141 ICARVDASFOPOGSLAPTAQRCPTY-----MDVIVLDGNSIYPMSEVO----- 185  
 DB 163 YCRSKGGS---TGEKKASARRRTVEEDEEDGTETALVDGSGSIEF--SDPGAKNFI 218  
 QY 186 -TFLRLVGLFLDPEQIOVGLVQGESPVHEWSLGDPR-----TKEEVRAAKNLSRR 238  
 DB 219 STMRNFYKCF---ECNFALVQGVAVIQTEFDLOESRDINASLAKVSIQVKEV--- 271  
 QY 239 EGRETQAQAIWACTEGFSQSHGRPEARLILVVTVDGESH--DGEELPALKACAGHV 297  
 DB 272 ----TTLTASAMQHVLDNITIPSRGSKKALKWVVLTDGIDGDPNLTLTVNSPRMOCV 327  
 QY 298 TRYGVIAVLGHLRORPSSFLNEIRTIASDPERFFNVVTDEALTDIVDALGRIEGL 357  
 DB 328 VAFALIGV--GDAPFN---NNTYRELKLIASDPKEATFVKTNSALDGLSLKLOQRIYAM 382  
 QY 358 ESSHANESFGLIENSQIGFSTRHLKDQ--ILRGMVATYGM--GGSVIM--LEGHRLPFRPM 414  
 DB 383 EGTVGD---ALQYQALQGFSAQIIDKQGVLLGTVAFAWMSGALYLSFONRGCELNQT 439  
 QY 415 ALDEDEPPALQNHAAVLGYSSVSMILRGRLLEFGAPRRHKGXTIAFOLKKDGAVRVA 474  
 DB 440 AKED-----SRTOYSLTGLSL--AVLKAHGISIVAAAPRHLKRGAV--TELKREDEEDA 492  
 QY 475 --QSLQGOIGSYFSGELCPLDTRDGTITDVLVAAFMELGPONKTEGRVYVYLVOQO--S 531  
 DB 493 FPRRIEGEMGYSFGSVLCPVIDMDGTTDILVAAFPY--HIRGEGRVYVYQVEBQA 550  
 QY 532 LTLTQCTLOPPEP--QDARFGFAMGALPDLNODGFALVAVGAPLE-----DG--HQAALYLY 584  
 DB 551 SFSLAHTLSGHPGLTNSRFGFMAAAGDINODKFTDVAIGAPLEGAGDAGSAGSVYIY 610  
 QY 585 HGTOGQVRHPAPORIAAAMPALSYFGRSVDRGLDLDGDDLVDAVVAQGAAILLSRPP 644  
 DB 611 NGHSGGLYSPSQIIRASSVASGLHFGMSVSGGLDFNPDGLADITVGRDAVILRSRP 670  
 QY 645 IVHLTFSLEVTPOAISV---ORDCRRQGEAVCLTALCFQVTSRT---FGRMDOHY 697  
 DB 671 VVDLTVSMFTPDALPMVFILGKMDV-----LCFEVDSVVAASEPQL--REWF 716  
 QY 698 MRFASLDWMTAGARAFDGSQORLSPRRLISVGNVTEQOLHFHYLDT-----SDLYRP 752  
 DB 717 LNFETVDVDTKORORLQCEBSSCCSCCLP--KNNGSFLCE--HFWLISIEELCEBECFEN 773  
 QY 753 VALTVTFALDNT---TKPGPVINE--GSPTSIQKLVFSGDCGDPNCECTDVLQVNMNI 807  
 DB 774 ITIKTYEFQTSGRRDYNPFTLDHAKESALFOL--PYEKCKNKKFCIAETQLTNNISQ 832  
 QY 808 RSRKAPFVVGGRKRVLTSTLENKENAVENTSIIISRLHLASLTPORESPIK--V 865  
 DB 833 QE-----LVGVTEKVTNNISLTNSGEDSYMTNMAALNPRNLQPKI-----QKPVSPV 882  
 QY 866 ECASASAHARL-----CSVHPVFO---TGAKYTFLLFEFFSGSLISQVFGKLTASSDGL 918  
 DB 883 QCDDEKPVASVIVMCKIGHPIIKRSVAVSVTWOLE-----ESVFPNRTADITVT 933

QY 919 ERNGTLOENTAOISAVIQYEBPHLLFSSESTLHREYVHPYGLPVGPGEFTLLRVQNLG 978  
 DB 934 ISNSN--EKSILARETSLQFRIHAFI-----AVLSRPV--MYMTSOSPSPHKEFFNV----- 983  
 QY 979 CVVSGLLIISALLPAVHAGNVFLSLQVITNNASCI-----VQNLTEPPGPV 1027  
 DB 984 -----HGENLFGAVPOL-----QICVPIKLQDFQIVRVKULTKQD--- 1019  
 QY 1028 HPEELQHTNRLNGSTQOQVVRCH-----LQGLAKGTESVSG-----LLR-----L 1068  
 DB 1020 HTECTOSEPACSDPVQHVKEHMSVCAITSNKENVVAABISVGHTKQLRDVSELPI 1079  
 QY 1069 VANEFFRAKFKSL-----TVSSTFELTEGVSVLQUTEASRMSSESLLEVOTRPIIS 1122  
 DB 1080 LGEISFNKSLVEGLNAENHRTKITVFLKEE-----TR-----S 1114  
 QY 1123 LMIIGSVLGGLLLALLVFCIMKLGFPANR--KIPBEKBEKLE 1166  
 DB 1115 LPLIGSSIGGLVVLVITAILFKCGFPKRYQQLNLESTRRAQLK 1160

RESULT 12  
 Q98TF0 PRELIMINARY; PRT; 1187 AA.  
 AC Q98TF0;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE CD11-2.  
 GN C1A2.  
 OS Cyprinus carpio (Common carp).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Cyprinus.  
 NC NCB1\_TaxID=7962;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PERITONEAL EXUDATE CELLS;  
 RA Kimura M., Fujiki K., Nakao M.;  
 RT "Molecular cloning of a leukocyte integrin from the common carp."  
 RL Submitted (Sep-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB048537; BAB39135.1; -  
 DR HSSP; P20701; ILPA.  
 DR InterPro; IPR001969; Asparticase site.  
 DR InterPro; IPR002035; vwf\_A.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR Pfam; PF01839; Fg-GAP; 5.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00193; Int\_alpha; 5.  
 DR SMART; SM00327; vwa; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 DR PROSITE; PS50234; vWFA; 1.  
 SQ SEQUENCE 1187 AA; 131778 MM; 85EDC7CA8B6B1C59 CRC64;

Query Match 15.5%; Score 947.5; DB 13; Length 1187;  
 Best Local Similarity 26.7%; Pred. No. 1.5e-62;  
 Matches 333; Conservative 209; Mismatches 504; Indels 199; Gaps 50;

QY 5 FVTH-----LPLVFLTGLCSFPLNDEHRLPFGPEAEFGYSVLQHVGGGGRMM 56  
 DB 6 FVTQSSRMETLGLILFMASLSBAFNIDTEHPLRFNGTEDEDFGYSVYOTEGNRKQI 65  
 QY 57 LVGAPMDPSSGDRGDVRCPEVGAHNAAPCAKGLGDIQLOANSHPAVNMHLGMSLLET 116  
 DB 66 IVGAPLENSA---GEMVSCF---ADLQSKR-----LGRPSESVRF--FGMSAASV- 110  
 QY 117 DGGGFMACAPLMSRACGSSVSSGICARVDASFOPOGSLAPTAQRC--PTYMDVIVLVDGS 175  
 DB 111 -SAALTSCSPYFAHECDGNSYLVNGVYQFNSLQAVSNFTAAVQECSSKREVNVLVLPDGS 169



Qy	622	DGDDLVDAVGAQGAAILLSRPIVHLTPSLEVTTPAIVSDRCRRGEAVCLTAALC	681
Db	647	SGDDLADIVGSDVAVALRSRPVVDLTYSMTPTPALMAFCD	697
Qy	662	FOVTSRT---DGRWDHQFYMRFTSLDMEWTAGAAAFDGSQGL	729
Db	698	FKVDSAAVPSSEBGL--RGMSLNFTVDV-----VTKOKORLCADARCCQSCCLMKK	746
Qy	730	SVGNATTCBGLHNVHLDT-----SDYLRVALVTPEALDNTK---PGPVUNE-GSPFS	778
Db	747	SGGSSLCE--HGGLSTEELCEDDCFSNITTVASIEFQTSBERRHNPVLDHYKEPS	804
Qy	779	IQKLVFSCDKCPDNECVYDVLVQVMMDIRGSRKAPFVVRGGRKVLVSTLENKRAY	838
Db	805	IFQL-PYEKDCGNKVFCEIAEIQLTTAISQOD-----LVGITKEVTNMNLSITNSGEDSY	857
Qy	839	NTSLSLITSRNLHLASLTPQRESPIKVECAAFSAHARL-----CSVGHVPVQ--TGAKYT	891
Db	858	MTNMLNPNRNQIFKKI--QKPLSPDIQDDPKPVASVLMWCKIGHPIILKSSVNVST	915
Qy	892	FLFEFEFCSSLSLQVFGKLTAS-----SDSLERNITLQENTPAQ-TSAYIQY--EPLILF	943
Db	916	WGLE-----ESIFPKRTADITYTINSNEKSLARETHSIQFPHALIVALSRSVWY	966
Qy	944	---SSESTLHRYEVHPDYGLTVPGVGEPEFTTLRVONLGVYVSGLIISALLPAVAHGNY	1000
Db	967	MNTSOSSSDHK-----EFPFNV-----HGENH	988
Qy	1001	FLSLSGVITNNASCTIVQNTPEPGPVHHELO-----HTNRLNSG--TQCCVVAHCHG	1050
Db	989	FGAVYQL---QICV-----PITRLDQITRVKHLTKQAHTECTOSQEPICSD	1030
Qy	1054	QIAKTEVSVGLRLVHNEFFRAKFKSLTVVSTFELG-----TEBGSVLQUTLEASRWE	1100
Db	1035	PVQNVHEMHVSICALTSNK-----ENVTVAAEISMGHTKQLRLDISELQIGEISFNK	1080
Qy	1109	SULEVY-----QTRPILI-----SLWITLGSVVGILLALVFCMLKIGFF--ARK	1150
Db	1088	SLYEELNENHHTKTVIFLKEEKPHSLPILITIGSSIGLVLVAILALFKCGFFRKRYK	1140
Qy	1154	KIPEEKEEKELE	1166
Db	1148	QLNLESVKKAQLK	1160
RESULT 14			
Q28984			
Q28984	PRELIMINARY;	PRT;	920 AA.
AC	Q28984;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	CD11b (Fragment).		
GN	CD11B.		
OS	Sus scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
NC	NCBI_taxonomy=9823;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RA	Lee J.-K., Schook L.B., Rutherford M.S.;		
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: U40072; AAB16869.1; -		
DR	HSSP; P11215; IABX.		
DR	InterPro; IPR000413; Integrin_alpha.		
DR	InterPro; IPR002035; VWF_A.		
DR	Pfam; PF01839; FG-GAP; 4.		
DR	Pfam; PF00092; vwa; 1.		
DR	PRINTS; PRO1185; INTEGRINA.		
DR	PRINTS; PRO0453; VWFADOMAIN.		
DR	SMART; SM00191; Int_alpha; 4.		
DR	SMART; SM00327; vwa; 1.		
DR	PROSITE; PS50234; VWFA; 1.		

FT	NON TER	1	1
SEQ	SEQUENCE	920 AA, 102440 MW, E96CC51E350DD5AC CRC64;	
QY	Query Match	14.5%; Score 883.5; DB 6; Length 920;	
	Best Local Similarity	29.1%; Pred. No. 6.9e-58;	
	Matches	288; Conservative 165; Mismatches 376; Indels 161; Gaps 39	
QY	157	PTAQR-CPTY-MDVYIVLFDGSSNSI--YMSVEQTFRLRLVGLFLDPEDIGVGLVQYGS	212
DB	5	PEALRGCCQOQSDLAFLIDGSGSIRLDFORMKEVSTVMGO--FOKSKTFLALMOYSD	62
QY	213	PVHEMSLGDPR--TKEEVRAAKNLSRREGREYTAQAIWVACTEFSQSHGPREAR	269
DB	63	FYTHPTFNDPFRNPSPKLLVPIRQLGR---THTATGIRKVRRELFFSKSGARENALK	118
QY	270	LLVYVTDSESHDETLPALAKACEGR--YTRYGLAV---LGHYLRQRDPSPFLREIT	324
DB	119	ILVYITDEKFR-GPFLGYEDVIPEADRKGVIIRYVGVDAFNSMKSRE-----ELMT	169
QY	325	IASDEDERFFPNVTDALTDIVDALGDRIFGLSESHAHENESSFGLMSQIGFSTRRLD	384
DB	170	IASKRCGHVQVANNFEAVKTIQNOLOEKTAIBETQGSITSFPCFMSQBEFSALITSN	229
QY	385	GILFGMGAYWM-GGSVIMLEGGHRLFPPRMALDEPPALONHAAYICYSVMILRCG	443
DB	230	GPLLGAVSPFMWAGAFLLHMPKDRVIFINTTRVDSM-----NNAVLYGAV-EVILRMQ	282
QY	444	RRLFLSGAPRRHKGKVIATFOLKKDGAARVAQSLGEOIGSYFSGSELCPLTDPRDQTTDV	503
DB	283	AOSLVLAGPRYQHTGCLVWFK-QNSGAKENKADIGSQISYFGSGLCSVDVNRDSSDL	341
QY	504	LLVAAPMFLG-----POKKEGRVYVVLVGOSSLTLLTGTLQEPEDPARFGF	551
DB	342	VLIAGPHYEQTRGGQVSVCELPQGRAKMOCRVILCGEG-----HPW-----SRFGA	389
QY	552	AMGALPRLNOCGFADVAVGAPLEDEGHQALVLYHGT-QSGVRPHPAQRILAASMHALSY	610
DB	390	ALTAAGDVNGKLTIDVALGAPGEONRRAYVLFHTISLGSISPHSQRAGLSPLRLOY	449
QY	611	FGRSVDGRLDGDDDLVAVAGAOAAALLSSRPVHLTPSLLEVPOAISVVOQDCRRRG	670
DB	450	FGGSLSGGQDILTMGLMDLAVGAQGHVFLNLSQPLRAVBSNVFRRREARNAVYECRQQA	509
QY	671	QEA-VCLTAALCFQVYTSKTPGRMDH-----QFYMKFPAUSDWMTAGAPAAFDGSGQRIS	723
DB	510	AKQTAGEVQVCLQYRKST--MPLREBDQTSIITYDLADPGRPHRAVFEET--KKN	564
QY	724	PRRLRLSAG-VWTCGQLHFVLVD-TSDYLRPALVTFALDWTTRKG-----PYLNG	774
DB	565	TRRQQTIGLSRKCEHLALMPDCYEDSVTPVLRINSL-VGRPASSFGULRYLVAWD	622
QY	775	SPTSIQKLVPSPKDCGPDNECVTDVLVQ--VNMDIRSGRAFYVGRGRVLYSTLE	831
DB	623	AQRLFTALFFPEKKGANDSICODDLISIFSFMSLDT-----LVVGGPRLKATLYLR	674
QY	832	NRKENAYTSSIIIFSRMLHLASLTPQR-----ESPIVECAPAPAHARLGSV	879
DB	675	NOGEBSYTYQVTFPYPSDLSTRKYSTONSQNSQSRMLACEBSVSTE-ESTALKSTGSI	733
QY	880	GHPVQOTAKYTLFLLEFFSCSLSLQVFGKLTASSDSLERGTLOENTAQ-----	930
DB	734	NHPIFPDQSEVTFPNTVPVNDPDLGY--KLILKANTSENNMPSNKTEFQLEPLVXY	790
QY	931	-----TSAYIQEBHLLFSSESTLARYEVHPYGLPVGPGPEFKTTLARVONLGCYVSG	984
DB	791	AVYVVVTSLEVSTKYFNFATSEKTRHYIE-HQY-----QFNMLG--QRK	831
QY	985	LIISAL-----LRVAHAGNYFLSLSGOVITNNASCIQVNLTEPPGPVHP--EELQRT	1033
DB	832	LPISVVEVNPVRLNRYVWMDQPVYTFSSONLSRSCS-----TEELIGFR-HSPFLKLOXT	884
QY	1036	NRLNGSNTOQCVNRCCHLQGLAKTEVSVGL	1065

```

Db      885 PVLNCSIAVCQKICQDIPSGIQEELKYL 914

RESULT 15
Q9WUF8      PRELIMINARY; PRT; 895 AA.
ID Q9WUF8
AC Q9WUF8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Itgae protein (Fragment).
GN ITGAE.
OS Mus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10095;
RN [1]
RP SEQUENCE FROM N.A.
RA Schoen M.P., Arya A., Murphy E.A., Adams C.M., Strauch U.G.,
RA Agace W.W., Marsal J., Donohue J.P., Her H., Beter D.R., Olson S.,
RA Lefrancois L., Brenner M.B., Grusby M.J., Parker C.M.;
RT "Mucosal T lymphocyte numbers are selectively reduced in integrin
RT alphae (CD103) deficient mice."
RL Submitted (Mar-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF133085; AAD30063.1; -.
DR EMBL; AF133070; AAD30063.1; JOINED.
DR EMBL; AF133071; AAD30063.1; JOINED.
DR EMBL; AF133072; AAD30063.1; JOINED.
DR EMBL; AF133073; AAD30063.1; JOINED.
DR EMBL; AF133074; AAD30063.1; JOINED.
DR EMBL; AF133075; AAD30063.1; JOINED.
DR EMBL; AF133076; AAD30063.1; JOINED.
DR EMBL; AF133077; AAD30063.1; JOINED.
DR EMBL; AF133078; AAD30063.1; JOINED.
DR EMBL; AF133079; AAD30063.1; JOINED.
DR EMBL; AF133080; AAD30063.1; JOINED.
DR EMBL; AF133081; AAD30063.1; JOINED.
DR EMBL; AF133082; AAD30063.1; JOINED.
DR EMBL; AF133083; AAD30063.1; JOINED.
DR EMBL; AF133084; AAD30063.1; JOINED.
DR HSSP; P11255; IABX.
DR InterPro; IPR004113; Integrin_alpha.
DR InterPro; IPR020315; VWF_A.
DR Pfam; PF01839; FG-GAP; 4.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00451; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWF_A; 1.
FT NON_TER
FT 1
FT 895
SQ SEQUENCE 895 AA; 98266 MW; BEEEA14A754ADAIE CRC64;

Query Match 14.5%; Score 883; DB 11; Length 895;
Best Local Similarity 30.5%; Pred. No. 7,2e-58;
Matches 295; Conservative 154; Mismatches 333; Indels 186; Gaps 48;

QY 23 ENLDEHH--PRLFGPPAEAFEGYVLOHV--GGGQRMVLVGAPWDGPGSGRRGDVYRCP 77
DB 7 FNMDDVMAMVTLQCGAPAV--LSSLHLDPSSNQTCLVAR---RSSNRNTALYRCA 60
QY 78 VG-GAHNAPCAR-GHL-----GDYOLGNSHRAVNMHLGMSLLETDDGCFMACPLMGR- 130
DB 61 ISISPEDEIACQPEVHEHCMPKGRYQ-----GVTLV--GNHNGVLVCIOVOAK 105
QY 131 -----AC-----GSSVFSSGICARVDASFQPGSLA 156
DB 106 FRLNSLSELTGACSLTLPNDLQAQAVFSDLBEGFLDPAHVDGDIYCRSKGS---TGEEK 162
QY 157 PTAQCPCFY-----MDVIVILDGNSIYPMSEVQ-----TFLRLVGLKPLIDPE 200

```

```

Db      163 KSARRRTVEEDEDGTEIAIYLDGSGSIEP-SDFOKANFISTMNRNFEKCF----- 217
QY 201 QIOVGLVOYGESPVHEWELGDPR-----TKEEVRAKANLSRRRGRETAKTAQAIMVACT 254
Db 218 ECFNALVOYGAIVQIEFPLQESRDINSLAKVQSIYQKEV-----TKTRASMOHVD 270
QY 255 EGFSGSHGRRPEARLVLVVTGDSEH-DGEELPAALKACEAGRYRYGIAVLGHYLROR 313
Db 271 NIFIPSRGSRKALKVMVTLTDGDFGDPPLNLTITVINSPPKQGVVRFAGV-GDAFKA-- 327
QY 314 DPSSFLREIRITIASPPDRFPFNVTDEALVDIYDALDDRIFGLEGSAENESSGLEMS 373
Db 328 --NMTYRELKLIASDPKKAHTFKYTNYSALDGLLSKQQRIVHMGIVGD--ALQYOLA 382
QY 374 QIGFSTHLKKG-IDFGWVAYDW-GSGVLM-LEGHRLFPFRMLDEFPALQNHAY 430
Db 383 QTFSAQLLDKGQVLGLVGAFPMNSGGLVISTONGRCFLNQTAKED---SRIVQYISY 438
QY 431 LGYSVSMILRGRRFLSGAPRRHRGKVIAFOLKDGAVRA--QSLQGEQISYFGS 488
Db 439 LGYSL-AVLHKAHGISYAGAPRHLRGAV--FELRKEDREDAFVRRIEGBQMGSYFGS 495
QY 489 ELCPDLTRDGTDTVLVAAMPFLCPQKKEGRVYVYLVGOQ-SLITLQGTIQEPP-QD 546
Db 496 VLPVDIDMDGDTTFLVAAPFY--HIRGEGRYVYVQVPQDASFLAHTLSGHPGLTN 553
QY 547 ARFGFAMGALDNLNDGFADVAVAGPLE----DG-HOGALYLYHGTSQGVRRPRAORIA 600
Db 554 SRFGPMAAVGDINDKKTDTVAIGPLEGFGAGDASIGSYIYINGHSGGLYDSFQQR 613
QY 601 AASMPHALSYFGRSYDGRULDGDLDVDAVAGCAAILSSRPVHLTPSLVTPPAIS 660
Db 614 ASSVASGLHYFGMSVSGGLDFNGDGLADITVGSRSAAVILSRPVDLTIVSGTFTPDALP 673
QY 661 VV---QPRCKRRGCEAVCLTALCFQVTSRT---PGRWDHOFWRFTASLDEMTAGARA 713
Db 674 MVFIQKMDVN-----LCFEVDSSVVASEPGL--REMPUNFTVDVDTVKORQL 719
QY 714 AFDGSGQRSLSPRLRLSYGVNTCEQLHPHVLDT-----SDYLRPVALTFTVFDLNT 764
Db 720 QCEDSSGQSCLR-KMNGSFLCE--HFWLISTELCEDEDFPSNITTITVTEFQSGARR 776
QY 765 TKPGFVLNE-GSPTSIQLVPEFSKDCGPDNCCVTDLVQVNMIDGRSKAPFVVRGGRK 823
Db 777 DYPNPTLDHYKEPSAIFQL-PYEKDCKKVPCIAIEIQLTINISQGE-----LVVGATKE 829
QY 824 VLVSTLTENRKENAYNTSLSIIFSRNLHLASLTLPQRESPIK--VECAAPSAHARL-----C 877
Db 830 VTMNISLTNSGEDSDSYMIMALNYPRLQFKKI-----QKRVSPDVQCDPDKPVASVLVNNC 885
QY 878 SVGHVPFQ 885
Db 886 KIHFILK 893

```

Search completed: July 16, 2003, 07:55:23  
Job time : 126.819 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:22 ; Search time 129.667 Seconds

(without alignments)  
1199.256 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106  
Sequence: 1 MELPVTHTLFLPLVFLTGLC.....GFFAHKKIPEEKREKLEQ 1167

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A.GeneSeq 101002.\*

```

1: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6106	100.0	1167	21	Human integrin sub
2	6060	98.9	1167	22	Human secreted pro
3	6031	98.8	1152	22	Human secreted pro
4	5965	97.7	1152	22	Human secreted pro
5	5904.5	96.7	1132	21	Human integrin sub
6	2439.5	40.0	1188	22	Human novel protei
7	2439.5	40.0	1188	22	Human A259. Homo
8	2439.5	40.0	1188	23	Human A259. Homo
9	2438.5	39.9	1188	22	Human A259. Homo
10	2429	39.8	1189	21	Human novel protei
					ITGA11 protein enc

11	2429	39.8	1189	22	ABG12949
12	2422	39.7	1188	22	AAAB50087
13	2422	39.7	1188	23	AAAB10552
14	2403.5	39.4	1188	22	AAAB30929
15	2254	36.9	1034	21	AAAB25590
16	1894.5	31.0	1180	23	ABBB90788
17	1893	31.0	1179	23	ABBB90759
18	1831	30.0	1183	20	AAV07728
19	1730	28.3	1183	20	AAV07729
20	1727.5	28.3	707	22	AAU19663
21	1727.5	28.3	707	23	ABP47883
22	1674	27.4	1367	19	AAW70542
23	1476	24.2	303	21	AAV32822
24	1363.5	22.3	979	22	ABG29239
25	1131.5	18.5	1161	16	AAW78166
26	1131.5	18.5	1161	18	AAW23049
27	1131.5	18.5	1161	19	AAW72825
28	1131.5	18.5	1161	19	AAW65089
29	1131.5	18.5	1161	19	AAW57491
30	1131.5	18.5	1161	20	AAW73342
31	1131.5	18.5	1161	21	AAW07359
32	1131.5	18.5	1161	23	ABG61468
33	1126	18.4	1161	18	AAW23064
34	1126	18.4	1161	19	AAW72837
35	1126	18.4	1161	19	AAW65106
36	1126	18.4	1161	20	AAW73343
37	1126	18.4	1161	21	AAW07376
38	1126	18.4	1161	23	ABG61485
39	1124.5	18.4	1161	16	AAW78169
40	1124.5	18.4	1161	18	AAW23062
41	1124.5	18.4	1161	19	AAW72824
42	1124.5	18.4	1161	19	AAW60004
43	1124.5	18.4	1161	21	AAW07374
44	1124.5	18.4	1161	23	ABG61483
45	1118.5	18.3	1161	19	AAW65104

## ALIGNMENTS

RESULT 1	AA32242	
ID	AAV32242	standard; Protein; 1167 AA.
XX	AAV32242;	
AC	AAV32242;	
DT	15-FEB-2000	(first entry)
XX		
DE	Human integrin subunit alpha-10.	
XX		
KW	Integrin alpha-10; ISAL0; human; trauma; rheumatoid arthritis;	
KW	osteoarthritis; osteoarthritis; cancer; atherosclerosis;	
KW	inflammation; therapy; cartilage; chondrocyte; osteoblast;	
KW	fibroblast; vaccine; marker.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..22
FT	Protein	/note= "signal peptide"
FT		23..1145
FT		/note= "mature protein"
FT	Domain	23..1120
FT		/note= "extracellular domain"
FT	Domain	1121..1145
FT		/note= "transmembrane domain"
FT		1122..1167
FT		/note= "cytoplasmic domain, specifically claimed
FT	Domain	in Claim 21"
FT		162..359
FT		/note= "I-domain"
FT	Binding-site	494..502
FT		/note= "cation binding site motif"

Novel human diagno  
Murine A259. Mus  
Murine A259. Polype  
Amino acid sequenc  
Protein encoded by  
Rat Tumour Endothe  
Human Tumour Endot  
Armenian hamster a  
Human hamster a  
Human novel extrac  
Human polypeptide  
Integrin alpha-2 c  
Mouse integrin sub  
Novel human diagno  
Human beta-2 integ  
Human beta-2 integ  
Human alpha-d. Ho  
Human beta-Integri  
Human beta2\_1 integr  
Human alpha\_d\_1. R  
Rat alpha\_d polype  
Rat alpha\_d protei  
Rat beta2\_integri  
Rat beta-integrin

FT Binding-site 558..566  
 FT Binding-site /note="cation binding site motif"  
 FT Binding-site 620..628  
 FT Binding-site /note="cation binding site motif"  
 FT Modified-site 98  
 FT Modified-site /note="N-glycosylated"  
 FT Modified-site 336  
 FT Modified-site /note="N-glycosylated"  
 FT Modified-site 364  
 FT Modified-site /note="N-glycosylated"  
 FT Modified-site 733  
 FT Modified-site /note="N-glycosylated"  
 FT Modified-site 839  
 FT Modified-site /note="N-glycosylated"  
 FT Modified-site 921  
 FT Modified-site /note="N-glycosylated"  
 FT Modified-site 1018  
 FT Modified-site /note="N-glycosylated"  
 FT Modified-site 1039  
 FT Modified-site /note="N-glycosylated"  
 FT Modified-site /note="N-glycosylated"  
 XX MO951639-A1.  
 XX  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 31-MAR-1999; 99WO-SE00544.  
 XX  
 PR 02-APR-1998; 98SE-0001164.  
 PR 28-JAN-1999; 99SE-0000319.  
 XX  
 PA (ACTI-) ACTIVE BIOTECH AB.  
 XX  
 PI Lundgren-Akerlund E;  
 DR MPI; 2000-052639/04.  
 XX  
 DR N-PSDB; AA234719.  
 XX  
 PT New isolated integrin subunit alpha-10, used as a marker or target  
 PT molecule for cells during development, regeneration and pathological  
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 PT inflammation -  
 XX  
 XX  
 XX Claim 1; Fig 6; 90pp; English.  
 XX  
 CC This sequence represents novel human chondrocyte integrin subunit  
 CC alpha-10 (Isa10). A splice variant is given in AA932243. The  
 CC invention relates to a recombinant or isolated integrin heterodimer  
 CC comprising the alpha10 subunit in association with subunit beta  
 CC (especially beta-1). The heterodimer and the subunit alpha-10 can  
 CC be used as markers or targets of all types of cells, e.g. of  
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used:  
 CC for treating pathological conditions involving Isa10, such as  
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
 CC for detecting the formation of cartilage during embryonal  
 CC development, physiological or therapeutic repair of cartilage,  
 CC or detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes; for selection and  
 CC analysis or for sorting, isolating or purification of chondrocytes  
 CC and for in vitro studies of differentiation of chondrocytes; and as  
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
 CC skeletal muscle or other tissues where adhesion impairs the function  
 CC of the tissue (all claimed). Isa10 binding entities can be used to  
 CC determine the differentiation-state of cells during embryonic  
 CC development, angiogenesis or development of cancer, in pathological  
 CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,  
 CC in tissue regeneration or in therapeutic and physiological repair  
 CC of cartilage (claimed). A vaccine comprising the integrin  
 CC heterodimer or subunit alpha-10 is also claimed. Isa10  
 CC polynucleotides, vectors, host cells and methods of producing  
 CC recombinant Isa10 are also claimed.  
 XX  
 SQ Sequence 1167 AA;

Query Match 100.0%; Score 6106; DB 21; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELPVYTHLFLPLVLTGLCSFNNIDHHPRLPFGPPPEAFEGSYLVQHYGGGQRMVGA 60  
 DB 1 MELPVYTHLFLPLVLTGLCSFNNIDHHPRLPFGPPPEAFEGSYLVQHYGGGQRMVGA 60  
 QY 61 PMDGSGRRGDVYCPVGGANAPCAKGLGDYQLGNSSHVAMNHGMSLLETDPDGG 120  
 DB 61 PMDGSGRRGDVYCPVGGANAPCAKGLGDYQLGNSSHVAMNHGMSLLETDPDGG 120  
 QY 121 FMACAPLWSRACSSVFSSGICARVDASFOGSLAPTAQRCPYMDVYIVLDGNSIYP 180  
 DB 121 FMACAPLWSRACSSVFSSGICARVDASFOGSLAPTAQRCPYMDVYIVLDGNSIYP 180  
 QY 181 WSEVQTFRLRLVKGKFLIDPEQIQVGLVOYGESPVHEMSLGDRTKEEVYRAKKNLSRREG 240  
 DB 181 WSEVQTFRLRLVKGKFLIDPEQIQVGLVOYGESPVHEMSLGDRTKEEVYRAKKNLSRREG 240  
 QY 241 RETKTAQAIMVACTEGFSQSHGCRPEAARLLVVYVDGESHDEELPAALKACEAGRVTXY 300  
 DB 241 RETKTAQAIMVACTEGFSQSHGCRPEAARLLVVYVDGESHDEELPAALKACEAGRVTXY 300  
 QY 301 GIAVLGHYLRQRDPSFLREIRITIASDPDERFPFNVTDEAALTDIVDALGDRIFGLEGS 360  
 DB 301 GIAVLGHYLRQRDPSFLREIRITIASDPDERFPFNVTDEAALTDIVDALGDRIFGLEGS 360  
 QY 361 HAENESSFGLMWSQIGFSTHRLKDGILFGMVGAYWGGSVLMLBEGHRLFPFRMALDEDF 420  
 DB 361 HAENESSFGLMWSQIGFSTHRLKDGILFGMVGAYWGGSVLMLBEGHRLFPFRMALDEDF 420  
 QY 421 PPALONHAAVYGYSSVSMILRGGRFLFGAGRFRRGRVIAFOUKKQGAVVAQSLQGE 480  
 DB 421 PPALONHAAVYGYSSVSMILRGGRFLFGAGRFRRGRVIAFOUKKQGAVVAQSLQGE 480  
 QY 481 QIGSYFGSELCPLDTRDQDGTDDVLLVAAPMFLGPONKETGRVYVYLVGQSSLLTLQGTLO 540  
 DB 481 QIGSYFGSELCPLDTRDQDGTDDVLLVAAPMFLGPONKETGRVYVYLVGQSSLLTLQGTLO 540  
 QY 541 PEPPDARFGFAMGALPDLDNDFADVAVGAPLEBCHQCALYLYHGTOSGVRRPHRAQRIA 600  
 DB 541 PEPPDARFGFAMGALPDLDNDFADVAVGAPLEBCHQCALYLYHGTOSGVRRPHRAQRIA 600  
 QY 601 AASMPHALSYFRRSDVGRDLDDGDDLVAVAGAAAILLSRPIVHLTPSLLEVPOAIS 660  
 DB 601 AASMPHALSYFRRSDVGRDLDDGDDLVAVAGAAAILLSRPIVHLTPSLLEVPOAIS 660  
 QY 661 VVQRDCRRRQGEAVCLTAALCFQVTSRTTPGRWDHGFYMRFTASLDEWTAGARAAFDGSGQ 720  
 DB 661 VVQRDCRRRQGEAVCLTAALCFQVTSRTTPGRWDHGFYMRFTASLDEWTAGARAAFDGSGQ 720  
 QY 721 RLSPPRLRLSVGNVTCEDLHFVLDTSYLRVALTVFALDNTTRKPGVINEGSPSTIQ 780  
 DB 721 RLSPPRLRLSVGNVTCEDLHFVLDTSYLRVALTVFALDNTTRKPGVINEGSPSTIQ 780  
 QY 781 KLVPSSKCGPNECVTDVLQVNMDDIRGRAPRVVGGRRKVLVSTLLENRKKNAYVT 840  
 DB 781 KLVPSSKCGPNECVTDVLQVNMDDIRGRAPRVVGGRRKVLVSTLLENRKKNAYVT 840  
 QY 841 SLSIIFSRNLHLASTLPQRESPIKVECAAPSAAHRLCSGVHVPFGAKVTFLEFEFSC 900  
 DB 841 SLSIIFSRNLHLASTLPQRESPIKVECAAPSAAHRLCSGVHVPFGAKVTFLEFEFSC 900  
 QY 901 SLSLVQNGKLTASSDSLERNCTLOENRQTSAYIQYEPHLLFSEESTLHREVPYGYL 960  
 DB 901 SLSLVQNGKLTASSDSLERNCTLOENRQTSAYIQYEPHLLFSEESTLHREVPYGYL 960  
 QY 961 PVGPPEPFKTLRYONLGGYVVSGLISALLPAVHAGNYFLSLQVTTNNASCIVONLT 1020  
 DB 961 PVGPPEPFKTLRYONLGGYVVSGLISALLPAVHAGNYFLSLQVTTNNASCIVONLT 1020  
 QY 1021 BPPGPPVHPEELOHTNRUNGSNVTCQVVRCHLGLAKGTEVSVGLRLVHNEFFRRARKE 1080



```
Db      1021 EPPGPVHPEELQHTNRLNGSNTQCQVVRCHLGQIAKGTESVGLRLVHNEFFRRAFK 1080
Oy      1081 SLTVSTFELTGEESVLQTEASRWSSELEVVQTRPILISMLIGSVLGLLTLAL 1140
Db      1081 SLTVSTFELTGEESVLQTEASRWSSELEVVQTRPILISMLIGSVLGLLTLAL 1140
Oy      1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167
Db      1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167

RESULT 2
AAB64584
ID      AAB64584 standard; Protein; 1167 AA.
XX
XX
AC      AAB64584;
XX
XX      22-MAR-2001 (first entry)
DE      Human secreted protein #37.
KW      Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW      antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
KW      vulnerrary; anticoagulant; antibacterial; antifungal; antiparasitic;
KW      cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW      neurological disease; infection; human; secreted protein.
XX
XX      Homo sapiens.
OS
XX      WO200077197-A1.
XX      21-DEC-2000.
XX      01-JUN-2000; 2000WO-US14934.
XX      11-JUN-1999; 99US-0138599.
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      (ROSE/) ROSEN C A.
XX
XX      Rosen CA, Ruben SM, Komatsoulis GA;
XX      WPI; 2001-032312/04.
XX      N-PSDB; AAF32793.
XX
XX      Isolated nucleic acid molecule encoding a human secreted protein is
XX      used in preventing, treating or ameliorating a medical condition -
XX
XX      Claim 11; Page 496-500; 558pp; English.
XX
XX      Sequences AAB64549-B64594 represent the amino acid sequences of 47
XX      human secreted proteins encoded by the genes AAF32757-F32803. The genes
XX      and proteins are useful for preventing, ameliorating or treating medical
XX      conditions, e.g. by protein or gene therapy. The genes are isolated from
XX      a range of human tissues disclosed in the specification. The nucleic
XX      acids, proteins, antibodies and (ant)agonists are useful in the
XX      diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX      ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX      marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX      (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX      haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
XX      Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX      colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX      (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX      epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX      and parasitic infections.
XX
XX      Sequence 1167 AA;

Query Match      98.9%; Score 6040; DB 22; Length 1167;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1155; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
```

```
Oy      1 MELPVTILFLPLVETLGLCSFNLDEHHPRLPFPPEAEFCYSVLQHVGGQRMWLYGA 60
Db      1 VELPVTILFLPLVETLGLCSFNLDEHHPRLPFPPEAEFCYSVLQHVGGQRMWLYGA 60
Oy      61 PWDGSGRRGRDGYRCPVGAHANAPCAKGIHGLDYVLGNSSHPAVNMHGLMSLLETDGGG 120
Db      61 PWDGSGRRGRDGYRCPVGAHANAPCAKGIHGLDYVLGNSSHPAVNMHGLMSLLETDGGG 120
Oy      121 FMACAPLMSRACGSSVFSSGICARVDASFOPGSLAPTAORCPYMDVIVYLDGNSNIYP 180
Db      121 FMACAPLMSRACGSSVFSSGICARVDASFOPGSLAPTAORCPYMDVIVYLDGNSNIYP 180
Oy      181 MSEVQTFRLRLVGLFIDPEQIYVGLVYGESPVHEWSLGFERTKEEVRAAKNLSRREG 240
Db      181 MSEVQTFRLRLVGLFIDPEQIYVGLVYGESPVHEWSLGFERTKEEVRAAKNLSRREG 240
Oy      241 RETKTAQAIMVACTGFSQSHGCRPEAARLLVVTDSHDEELPALKACEAGRVTYR 300
Db      241 RETKTAQAIMVACTGFSQSHGCRPEAARLLVVTDSHDEELPALKACEAGRVTYR 300
Oy      301 GIAVLGHYLRQRDSSFLREIRTIASDPDERFFNVTDGALTDIYDALGRIFGLESS 360
Db      301 GIAVLGHYLRQRDSSFLREIRTIASDPDERFFNVTDGALTDIYDALGRIFGLESS 360
Oy      361 HAENESSFGLMSQIGFSTHRLKDGILFGMYGAYDMGGSVLMLEGHRLFPPRMALDEDF 420
Db      361 HAENESSFGLMSQIGFSTHRLKDGILFGMYGAYDMGGSVLMLEGHRLFPPRMALDEDF 420
Oy      421 PPALQNHAAVLYGYSSMLNRGRRPLISGAPRFRHKGVIAPQLKQGANVAOSLQGE 480
Db      421 PPALQNHAAVLYGYSSMLNRGRRPLISGAPRFRHKGVIAPQLKQGANVAOSLQGE 480
Oy      481 QIGSYFSGELCPIDRDGTTDVLVAAPMELGPNKGTGRVYVVLVGGQSLTLTGCTIO 540
Db      481 QIGSYFSGELCPIDRDGTTDVLVAAPMELGPNKGTGRVYVVLVGGQSLTLTGCTIO 540
Oy      541 PEPQDARFGFAMGALPDINODGFADVAVGAPLEDGHOALYLYHGTOSGYVPHPAORIA 600
Db      541 PEPQDARFGFAMGALPDINODGFADVAVGAPLEDGHOALYLYHGTOSGYVPHPAORIA 600
Oy      601 AASMPHALSYFGRSYDGRLLDGDLDLVVAACAQAAIILLSRPVHLTPSLEVPQAIS 660
Db      601 AASMPHALSYFGRSYDGRLLDGDLDLVVAACAQAAIILLSRPVHLTPSLEVPQAIS 660
Oy      661 VWRDCRRRGQAVCLTAALCFQVTSRTPGRWDHGFYMRFTASLDEMTAGAAAFDGGQ 720
Db      661 VWRDCRRRGQAVCLTAALCFQVTSRTPGRWDHGFYMRFTASLDEMTAGAAAFDGGQ 720
Oy      721 RLSPRRLSVGNVTCQQLHFFHVLDTSDYLRLPVALTVTFALDNTTKPGVNLNPGSPSTIQ 780
Db      721 RLSPRRLSVGNVTCQQLHFFHVLDTSDYLRLPVALTVTFALDNTTKPGVNLNPGSPSTIQ 780
Oy      781 KLVPSSKDCGPNNECVTLVLQVNDIGSRKAPFVVRGRRKVLVSTLLENKENAYNT 840
Db      781 KLVPSSKDCGPNNECVTLVLQVNDIGSRKAPFVVRGRRKVLVSTLLENKENAYNT 840
Oy      841 SLSLTFSNHILASITPQRESPIKVECAAPSAHNLCSVGHVPOTGAKVTTLFFPESC 900
Db      841 SLSLTFSNHILASITPQRESPIKVECAAPSAHNLCSVGHVPOTGAKVTTLFFPESC 900
Oy      901 SLSLQVGVKLTASSDSLERNGTLOENTAOISAVIYQEPHLLFSESTLHREVPYGL 960
Db      901 SLSLQVGVKLTASSDSLERNGTLOENTAOISAVIYQEPHLLFSESTLHREVPYGL 960
Oy      961 PVGPGPEFKTLRYONLCYVVSGLIISALLPAVAGGNYFLSLQVITNNASCIVONLT 1020
Db      961 PVGPGPEFKTLRYONLCYVVSGLIISALLPAVAGGNYFLSLQVITNNASCIVONLT 1020
Oy      1021 EPPGPVHPEELQHTNRLNGSNTQCQVVRCHLGQIAKGTESVGLRLVHNEFFRRAFK 1080
Db      1021 EPPGPVHPEELQHTNRLNGSNTQCQVVRCHLGQIAKGTESVGLRLVHNEFFRRAFK 1080
```

Oy	1081	SLTAVSTFELGTEESVQLTASRWSSESLSEVOTRPLISLWILISVIGGLLALL	1143
Db	1081	SLTAVSTFELGTEESVQLTASRWSSESLSEVOTRPLISLWILISVIGGLLALL	1140
Oy	1141	VFCLWKLGFPAHKKIPEEKREKLEQ	1167
Db	1141	VFCLWKLGFPAHKKIPEEKREKLEQ	1167
RESULT 3			
AAB64657			
ID	AAB64657	standard; Protein; 1152 AA.	
XX			
AC	AAB64657;		
XX			
DT	22-MAR-2001	(first entry)	
XX			
DE	Human secreted protein BLAST search protein SEQ ID NO: 167.		
XX			
KW	Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;		
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;		
KW	vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;		
KW	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;		
KW	neurological disease; infection; human; secreted protein.		
OS	Homo sapiens.		
XX			
PN	WO200077197-A1.		
XX			
PD	21-DEC-2000.		
XX			
PF	01-JUN-2000; 2000WO-US14934.		
XX			
PR	11-JUN-1999; 99US-0138599.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(ROSE/) ROSEN C A.		
XX			
PI	Rosen CA, Ruben SM, Komatsoulis GA;		
XX			
DR	WPI; 2001-032312/04.		
XX			
PT	Isolated nucleic acid molecule encoding a human secreted protein is		
PT	used in preventing, treating or ameliorating a medical condition -		
XX			
PS	Disclosure; Page 543-546; 558BP; English.		
XX			
CC	The invention relates to the isolation of genes AAP32757-F32803 encoding		
CC	the human secreted proteins AAB64549-B64594. The sequence is a search		
CC	result from a BLASTX homology search. The genes and proteins are useful		
CC	for preventing, ameliorating or treating medical conditions, e.g. by		
CC	protein or gene therapy. The genes are isolated from a range of human		
CC	tissues disclosed in the specification. The nucleic acids, proteins,		
CC	antibodies and (ant)agonists are useful in the diagnosis, treatment		
CC	and prevention of: (a) cancer, e.g. breast and ovarian cancer, and		
CC	other cancers of the adrenal gland, bone, bone marrow, breast,		
CC	gastrointestinal tract, liver, lung, or urogenital; (b) immune		
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic		
CC	anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,		
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)		
CC	cardiovascular disorders such as myocardial ischaemias; (d) wound		
CC	healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;		
CC	and (f) infectious diseases such as viral, bacterial, fungal and		
CC	parasitic infections.		
XX			
XX			
SO	Sequence 1152 AA;		
Query Match	98.8%;	Score 6031;	DB 22; length 1152;
Best local Similarity	100.0%;	Pred. No. 0;	
Matches 1152;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Oy	1	MELPFVTHFLPLVFLTGLCSPFNIDHHPRLLFGPPPAERGVSVLQGVGGQGRMWLIGA	60

Dd	1	MELEPVTHTLFLPLVFTLGTCLSPFNLDEHHNRLFCGPPEAEAGYSVLQHVGGGQKMWLVGA	60
Qy	61	PWDEPSGDRDNDVYRCVPVGAHNAPCAKGHLGDOVLQNSSHPANMHLGMSLLETDDGG	120
Dd	61	PWDEPSGDRDNDVYRCVPVGAHNAPCAKGHLGDOVLQNSSHPANMHLGMSLLETDDGG	120
Qy	121	FMACAPLMSRACGSSVSSGICAVDASFOQGSIAFTAORCPTVMVVLVDOSNIYP	180
Dd	121	FMACAPLMSRACGSSVSSGICAVDASFOQGSIAFTAORCPTVMVVLVDOSNIYP	180
Qy	181	MSEVQTLRLRLVGLFTIDPEQIOGLVQGESPVHEMSLGDFTKKEVYRAAKKLSRREG	240
Dd	181	MSEVQTLRLRLVGLFTIDPEQIOGLVQGESPVHEMSLGDFTKKEVYRAAKKLSRREG	240
Qy	241	RETTAOAIWACTEGFSOSHGGPEARLLVVTDESHDGEELPALKACEAGRTRY	300
Dd	241	RETTAOAIWACTEGFSOSHGGPEARLLVVTDESHDGEELPALKACEAGRTRY	300
Qy	301	GIAYLGHYLRORDPSSFLREIRTIASDPDERFFPNVTDEAALTIDALGDRIFGLEGS	360
Dd	301	GIAYLGHYLRORDPSSFLREIRTIASDPDERFFPNVTDEAALTIDALGDRIFGLEGS	360
Qy	361	HAHNESSFGLEMSQIGFSTRHKQGLFCMWGAIDMGSGVYLMEGHRLLPPRPAALDEF	420
Dd	361	HAHNESSFGLEMSQIGFSTRHKQGLFCMWGAIDMGSGVYLMEGHRLLPPRPAALDEF	420
Qy	421	PPALQNHAAALYLGYSVSSMLRGGRLEFLSGAPRRHKGKIAFOLKDKGAVRAQSLQGE	480
Dd	421	PPALQNHAAALYLGYSVSSMLRGGRLEFLSGAPRRHKGKIAFOLKDKGAVRAQSLQGE	480
Qy	481	QIGSYFSGELCPDLTDRDGTTLVLYAAPMFLGQNKETGRVYLYLVGQOSLTLQGTLO	540
Dd	481	QIGSYFSGELCPDLTDRDGTTLVLYAAPMFLGQNKETGRVYLYLVGQOSLTLQGTLO	540
Qy	541	PEPQODARFGAMALPDLNODGADVAVAGAPLEDHQGALYLHGQSGVRHPAPORIA	600
Dd	541	PEPQODARFGAMALPDLNODGADVAVAGAPLEDHQGALYLHGQSGVRHPAPORIA	600
Qy	601	AASMPHALSYFGRSVDGRLLDDGDDLVDAVAGAAILLSSRPVHLTPSLEVTPOAIS	660
Dd	601	AASMPHALSYFGRSVDGRLLDDGDDLVDAVAGAAILLSSRPVHLTPSLEVTPOAIS	660
Qy	661	VVGQDCCRRGGEAUCTALACFOYTSTTPGRMHOQFMRFLASIDEMTAGARAPDSGQ	720
Dd	661	VVGQDCCRRGGEAUCTALACFOYTSTTPGRMHOQFMRFLASIDEMTAGARAPDSGQ	720
Qy	721	RLSRRLRLISGANTCEQLHFHVLDTSIDYLRPVALTVPALDNTTKGPVLENGSPSIO	780
Dd	721	RLSRRLRLISGANTCEQLHFHVLDTSIDYLRPVALTVPALDNTTKGPVLENGSPSIO	780
Qy	781	KLVPFSKDCGPDNECVTDVLQVNMIDRGSRKAPFVTRGGRKVLVSTTLENKEMANYT	840
Dd	781	KLVPFSKDCGPDNECVTDVLQVNMIDRGSRKAPFVTRGGRKVLVSTTLENKEMANYT	840
Qy	841	SLSTIFSRNLHLAALTPORESPIVVECAASPAHARLCSVGPVQOTAKKTYLLEFEFSC	900
Dd	841	SLSTIFSRNLHLAALTPORESPIVVECAASPAHARLCSVGPVQOTAKKTYLLEFEFSC	900
Qy	901	SSLSIOVFGKLTASDSILERNGTLOENTAOQSAVIOYEPHLLFSSBSTLHRYEHPGTL	960
Dd	901	SSLSIOVFGKLTASDSILERNGTLOENTAOQSAVIOYEPHLLFSSBSTLHRYEHPGTL	960
Qy	961	PVGGEPEKTLRLRONIGCYVSSLLIISALLPAVAHGNYFLSIOYITNNASCTVONLT	1020
Dd	961	PVGGEPEKTLRLRONIGCYVSSLLIISALLPAVAHGNYFLSIOYITNNASCTVONLT	1020
Qy	1021	EPPEPVAHPBELQNTNRNLNGSNTQCVVRRCHLQGLAGTEVSVGLLVLHNEFFRRAKFK	1080
Dd	1021	EPPEPVAHPBELQNTNRNLNGSNTQCVVRRCHLQGLAGTEVSVGLLVLHNEFFRRAKFK	1080
Qy	1081	SLIVASTFELGTEEGSVLOLTEASRMESLLEVQTPRILISMTLIGSVLGGLLALL	1140
Dd	1081	SLIVASTFELGTEEGSVLOLTEASRMESLLEVQTPRILISMTLIGSVLGGLLALL	1140



Db 1141 VFCLMKLGFFAH 1152

RESULT 5  
AAV32243

ID AAV32243 standard; Protein; 1132 AA.

AC AAV32243;

DT 15-FEB-2000 (first entry)

DE Human integrin subunit alpha-10 splice variant.

XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;

KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;

KW inflammation; therapy; cartilage; chondrocyte; osteoblast;

XX fibroblast; vaccine; marker; splice variant.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..22 /note="signal peptide"

FT Protein 23..1132 /note="mature protein"

PN MO9951639-A1.

PD 14-OCT-1999.

XX 31-MAR-1999; 99WO-SE00544.

PR 02-APR-1998; 98SE-0001164.

XX 28-JUN-1999; 99SE-0000319.

PA (ACT1-) ACTIVE BIOTECH AB.

XX Lundgren-Akerlund E;

PI MPI; 2000-052639/04.

DR N-PSDB; AA234720.

XX New isolated integrin subunit alpha-10, used as a marker or target

PT molecule for cells during development, regeneration and pathological

PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or

PT inflammation

XX Claim 1; Page 43-48; 90pp; English.

XX This sequence represents a splice variant of novel human

CC chondrocyte integrin subunit alpha-10 (ISa10). It is identical to

CC ISa10 (see AAV32242) except for deletion of amino acids 975-986. The

CC invention relates to a recombinant or isolated integrin heterodimer

CC (consisting of beta-1). The heterodimer, subunit alpha-10 or splice

CC variant can be used as a marker or target of all types of cells, e.g.

CC of chondrocytes, osteoblasts and fibroblasts. They can also be used:

CC for treating pathological conditions involving ISa10, such as

CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;

CC for detecting the formation of cartilage during embryonal

CC development, physiological or therapeutic repair of cartilage,

CC of cartilage (claimed). A vaccine comprising the integrin

CC heterodimer or subunit alpha-10 is also claimed. ISa10

CC polynucleotides, vectors, host cells and methods of producing

CC recombinant ISa10 are also claimed.

XX Sequence 1132 AA;

XX Query Match 96.7%; Score 5904.5; DB 21; Length 1132;

XX Best local similarity 97.0%; Pred. No. 0;

XX Matches 1132; Conservative 0; Mismatches 0; Indels 35; Gaps 1;

XX

QY 1 MELPVTHFLPLVLTGLCSPPNLDHPRLPFPPPPAEFGSYLVQVGGGQRMVLYGA 60

DB 1 MELPVTHFLPLVLTGLCSPPNLDHPRLPFPPPPAEFGSYLVQVGGGQRMVLYGA 60

QY 61 PWDGSGRRGGVYVCPVGAHNAACAGHIGDYLGNSHPVAMNHGMSLLETDGGG 120

DB 61 PWDGSGRRGGVYVCPVGAHNAACAGHIGDYLGNSHPVAMNHGMSLLETDGGG 120

QY 121 FWACAPLMSRACGSSVFSSGICARVDASFOPGSLAPPAORCPTYMDVIVLDGNSIYP 180

DB 121 FWACAPLMSRACGSSVFSSGICARVDASFOPGSLAPPAORCPTYMDVIVLDGNSIYP 180

QY 181 WSEVOTFLRLVYKLFIDPEQIQVGLVOYGSSPVHMSLGDPRYKEEVVRAAKNLSREG 240

DB 181 WSEVOTFLRLVYKLFIDPEQIQVGLVOYGSSPVHMSLGDPRYKEEVVRAAKNLSREG 240

QY 241 RETKTAQAIMVACTGFSOSHGRPEARLLVVTVDGSHDGEELPALKKCEAGRVRY 300

DB 241 RETKTAQAIMVACTGFSOSHGRPEARLLVVTVDGSHDGEELPALKKCEAGRVRY 300

QY 301 GIAVLGHYLRQRDSSFLREIRTIASDPDERFFPNVDEALTDIVDAGRIFGLEG 360

DB 301 GIAVLGHYLRQRDSSFLREIRTIASDPDERFFPNVDEALTDIVDAGRIFGLEG 360

QY 361 HAENSSFGLEMSQIGFSTHRLKQILFGMGAYDMGSSVLMLEGHRLFPFRNALDEDF 420

DB 361 HAENSSFGLEMSQIGFSTHRLKQILFGMGAYDMGSSVLMLEGHRLFPFRNALDEDF 420

QY 421 PPALONHAAYIGYSSMLRGRRFLPSGA PRPHRGKVIAPOLKKGAVRAVQSLGE 480

DB 421 PPALONHAAYIGYSSMLRGRRFLPSGA PRPHRGKVIAPOLKKGAVRAVQSLGE 480

QY 481 QIGSYFGESELPIDDRGTTDVLVAAPMFLGPONKTEGRVYVYLVGQSLTLTQGTIQ 540

DB 481 QIGSYFGESELPIDDRGTTDVLVAAPMFLGPONKTEGRVYVYLVGQSLTLTQGTIQ 540

QY 541 PEPPODAFFGAMGALPDLNODGFADVAVGAPLEBGHOGALYLVHGTSQVRRPRAQRIA 600

DB 541 PEPPODAFFGAMGALPDLNODGFADVAVGAPLEBGHOGALYLVHGTSQVRRPRAQRIA 600

QY 601 AASMHALSYGRSDVGRDLDDGDLVAVAGAGAAIILSSRPVHLTPSLEVPQAI 660

DB 601 AASMHALSYGRSDVGRDLDDGDLVAVAGAGAAIILSSRPVHLTPSLEVPQAI 660

QY 661 VVQRCRRRGGAEVLTALCFQVTSRTPGRWDFQYVFTASLDEMTAGARAADGSGQ 720

DB 661 VVQRCRRRGGAEVLTALCFQVTSRTPGRWDFQYVFTASLDEMTAGARAADGSGQ 720

QY 721 RLSPPRLSLVGNVTCEDLHFRVLDTSYLARPAVLTTPALDNTTKPGPVINEGSP7SIQ 780

DB 721 RLSPPRLSLVGNVTCEDLHFRVLDTSYLARPAVLTTPALDNTTKPGPVINEGSP7SIQ 780

QY 781 KLVPSSKCGPDNECVTDVLQVNMDDIGSKRAPVVGGRKVIYSTLLENKRNANVT 840

DB 781 KLVPSSKCGPDNECVTDVLQVNMDDIGSKRAPVVGGRKVIYSTLLENKRNANVT 840

QY 841 SLSTIFSNLHLASLTPRESPIKVECAPAPAHARLCVSGHPVOTGAKVTFLEFEFSC 900

DB 841 SLSTIFSNLHLASLTPRESPIKVECAPAPAHARLCVSGHPVOTGAKVTFLEFEFSC 900

QY 901 SLLSIVFEGKLTASSDSLERNGTQENTAGTSAYIOYEPHLLPSSSESTLHRYEVPYGT 960

DB 901 SLLSIVFEGKLTASSDSLERNGTQENTAGTSAYIOYEPHLLPSSSESTLHRYEVPYGT 960

Db 901 SLLSQVRFKLTASSSLERNGLDENTQTSAYIOYEPHLLFSSSTLHREYHMYGTL 960  
 Qy 961 PVGPGEFEKTLRVONLGCYVVSGLIISALLPAVHAGNVPFLISQVITMNASCIYONLT 1020  
 Db 961 PVGPGEFEKTLR-----TNNASCIYONLT 985  
 Qy 1021 EPPGPVHPEEELQHTNRLNGSNTQCCVVRCHLGLAKGTEVSGILRLVHNEFFRRAKK 1080  
 Db 986 EPPGPVHPEEELQHTNRLNGSNTQCCVVRCHLGLAKGTEVSGILRLVHNEFFRRAKK 1045  
 Qy 1081 SLTVSTPELSTEGSGVQLTETASRWSSESLLEVONRPLISLMLIGSVLGLLTLAL 1140  
 Db 1046 SLTVSTPELSTEGSGVQLTETASRWSSESLLEVONRPLISLMLIGSVLGLLTLAL 1105  
 Qy 1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167  
 Db 1106 VFCLWKLGFPAHKKIPEEKREKLEQ 1132  
 RESULT 6  
 AAU14231  
 ID AAU14231 standard; Protein; 1188 AA.  
 AC AAU14231;  
 DT 24-OCT-2001 (first entry)  
 DE Human novel protein #102.  
 XX  
 KM Human; novel protein. Antianemic; osteopathic; antiinflammatory;  
 immunomodulatory; cytosolic; neuroprotective; vulnery; nocotropic;  
 anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;  
 thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 tissue regeneration; immune disorder.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155437-A2.  
 XX  
 PD 02-AUG-2001.  
 PF 25-JAN-2001; 2001WO-US02623.  
 PR 25-JAN-2000; 2000US-0491404.  
 PA (HISE-) HISEQ INC.  
 PI Tang YT, Liu C, Drmanac RT;  
 DR WPI: 2001-451939/48.  
 DR N-PSDB: AAS22536.  
 XX  
 PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
 PT nervous system disorders, and for regenerating bone and cartilage -  
 XX  
 PS Example 4; Page 578-581; 894pp; English.  
 XX  
 CC The invention relates to polynucleotides encoding novel human  
 CC proteins or their active domains. The polypeptides, polynucleotides and  
 CC antibodies raised against the polypeptides are used in a method of  
 CC treatment of a mammal and prevention of disorders caused by the aberrant  
 CC protein expression or activity. The polypeptides can be used as  
 CC molecular weight markers, food supplements, and in antibody production.  
 CC The polypeptides are used to identify compounds which bind to the  
 CC polypeptides. Polynucleotides of the invention are used as probes and  
 CC primers, for sequencing, for chromosome or gene mapping, in the  
 CC production of recombinant proteins, and in generating anti-sense DNA or  
 CC RNA and in gene therapy. Polypeptides of the invention can be used to  
 CC target drugs to a tumour, in assays to determine biological activity, to  
 CC raise antibodies/elicite an immune response, to determine quantitative  
 CC protein levels, as tissue markers, and to isolate receptors or ligands.

CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
 CC anti-inflammatory diseases, nervous system disorders, and infection.  
 CC The present sequence represents a protein of the invention.  
 XX  
 SQ Sequence 1188 AA;  
 Query Match 40.0%; Score 2439.5; DB 22; Length 1188;  
 Best Local Similarity 43.2%; Pred. No. 7.5e-208;  
 Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;  
 Qy 1 MELPFTVHLPLPLVEFLTGLCSFNLDEHHPLPFGPPAEFGYSVLQHVGGQRMVLVGA 60  
 Db 1 MDLPRLVAVAMALSLMPGFTDTPNMDTRKPRVIFGSRFAFGYTVQQHDISGNKVLVYGA 60  
 Qy 61 PWDGSGRRGDVYRCFPGAHNAPCAAGHLGDVQLGSSHPAYVMHLGMSLLETDCGG 120  
 Db 61 PLETNQYQKTGDVYKCPV---IHGCTKLNLGRVTLNVSVSRKDMRGLSLATNPKNDS 117  
 Qy 121 FMACAPLMSRAGSSVFSGGICARVDASFQGSAPPAOCPTMYADVIVLDGSSNTYP 180  
 Db 118 FLACSPPLMSHCGSSYYTTGCMCRVNSNFRSKTVAPLQRCQTYMDVIVLDGSSNTYP 177  
 Qy 181 WSEVOTFLRLRLVGLKFLIDPEOIQVGLVOYGSFVHWSLGDFTKEEVYRAKAKNLSREG 240  
 Db 178 WVEVQHFLINILKKRYIPGOIQVGVQYGDVHVEPLNDYRSVKDVAAASHIEQGG 237  
 Qy 241 RETKTAQIMACTGFSQSHGGRPEARLLVYVTDGSSHGEEPLPALKACEARVTRY 300  
 Db 238 TETRYAFGIEPARSBAFK--GGRGAKKVMIVITDGSHSDSPLLEKVIQOSEERNVTRY 295  
 Qy 301 GIAVLGHYLRORDDSSFLREIRITASDPDERFPFNVDAAALNDVIALDGRILGLEGS 360  
 Db 296 AVAYLVGYNRRGINDETFLNELKIYASDPDKHFNNVTDEALKDIVALGRILFSLGT 355  
 Qy 361 HAENESSFLEMSQIGFSTHRLKDGILFGWVAYDMGSSVLMLEGHRLFPPRMALDEDF 420  
 Db 356 N-KNETSFGLEMSQIGFSSHVVEDGVILGAAGAYDMNNAVLKESAGKAVILRBSYLKEF 414  
 Qy 421 PPALONHAAYIGYSVSMILRGRRRLFLSGAPRRPHRGKVIAPOLKCDGAVVAOSLQGE 480  
 Db 415 PEBLKNHAGVLYGVTVSVSSROGVVYVAGAPRFNHTKVLFTMHNNRSLTIHQAMRGQ 474  
 Qy 481 QIGSYFSGELCPDLPDRDGTDDVLLVAAWMLGQNKETGRVYVYLVGQGSLLTIOGTLQ 540  
 Db 475 QIGSYFSGELTSDVDGDGVTDLVVGAMFMN-EGRRERGVYVEL-RQNRFPVYNGTLK 532  
 Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADYAVGAPLEDHGQALYLYHGTQSGVRPHPAORI 599  
 Db 533 DSHSQYNARFSSSIASVADLNQDSYNDVVGAPLEDHNAAGIYIFHGRGSLIKTPKRI 592  
 Qy 600 AAASHPHALSYFGNSVSDGRDLDDGDLVDVAVAGQAAIILSSRPVHLTSLSEVTPQAI 659  
 Db 593 TASELATLQYFGGSIHQDLNEDGLDLVAGVILGNVILMSPPVVOINSLHPEPSKI 652  
 Qy 660 SVVORDCRRGOEAVCLTALACPQVTSRTPGRMWHQFMRFTASLDEWTAAGARAFDSSG 719  
 Db 653 NIFHDCRSRGRDTCFLAFLCTPPIFLAPFQTTVGIRINATWDERRYPRHALDGG 712  
 Qy 720 QRLSPRLRLSGVNVTCQLHFLVLDTSYDVRPALVTFALDNTTKGPVILNEGSPSI 779  
 Db 713 DRFNRAVLSSGGELGERINFHVLDTADYKVPVTFSEVYLEDPPH-GPMLDGGWPTTL 771  
 Qy 780 QKLVPSKDCGPPDNECVTDLVLQVNMDI-----RGRK-----APPV 817  
 Db 772 RVSVPFMNGCNEDHCVPDLVLDARSDLPTAMEYCOHRLRKPADCSAYTSLPDTVFI 831







QY	61	PMDGSGRRGRGVNYC	CPVGAAMNACCAQHLHDYOLGNSSHPAVNMHJGMSLFTDGGGQ	120
Db	61	PLETNGYOKTGDVYCPV	--IHGCTKLINLGRVTLNSVSEKDMNRJGLSLATPKNXS	117
QY	121	FMACAPLMSRACGSSV	FSFGICARVDASFQDGSJLAPTAQRCPYMDVYVILDGNSIYP	180
Db	118	FLACSPILMSHECGSS	YTTTGMCSRVNSNRFSGTVAPALORQOTYMDIYIVILDGNSIYP	177
QY	181	MSEVOTFLRLVGLKFL	FIDPEQIQVGLVOYGSEPYHEMSJGDPRTYEEVYVRAAKNLSREG	240
Db	178	WVEVOHFLINILTKKEY	IGBGOIQVENVVOYGEDVHAEFLINDRSYKDVAAESHLEORG	237
QY	241	RETKYAOALIMVACTG	EFQSHGSGREARLIVVWVWDGSHDEELPALKKACEAGVRVY	300
Db	238	TETRTAFGLIEFRASFA	FOK--GGRGAKKXMTVIDGSHSDPLEKXIQQOSEBNVTRY	295
QY	301	GIAYLVHYLRORDDSE	FLREIRITIASDPDERFENVVTDEALTDIVDGLRIJGLEGS	360
Db	296	AAVALGYNNRGINRET	FLNELKIYASDPDDKHFNVTDEALKDIVDGLGRIFSLGOT	355
QY	361	HAENESSGLEMSTQ	GFSTHRLKDGILGTMGAYIWMGSSVYMLBEGHRLFPFRALBDEF	420
Db	356	N-KNETSGLEMSQTF	SGFSSHVHVEDVLLGAGAYOMNGAVILEKTSAGVITPLRESYLFKEF	414
QY	421	PPALONHAAYIGYSSM	LLRCGRRLPLISGAPRFRHGRVIAFOLKXOAGVAVOJSGE	480
Db	415	PEELNKHAGVILGYIT	SVSSROGRVYVYAGARFNHTGTVILFTMHNNRSLTIHOAMGQ	474
QY	481	QIGSYFSGELCP	LDLDRDGTDTVLLVAAPMFLGPONKEGRVYVVLVGOQSILTLGOTIQ	540
Db	475	QIGSYFSGEITSV	IDGQVTDVLLVGAAPMYN--BGRGKXYVEL--RQRFPVYNGTLX	532
QY	541	PEPP--QDARFEGAMGAL	PDJLNDGADVAVGAPELDGHOGLYLYHGTQSGYRPHAPORI	599
Db	533	DSHSYQNAARFGSSI	ASVRLNDQSNVDVVGAPRLDNAGAIYIHFGRGSSILKTPKORI	592
QY	600	AAASYPHALSYFGR	SVDRJLDIDGQDIVVANGAAGAILLSRPIYHULTSLLEVTPAI	659
Db	593	TASELATLQYFGCS	IHQGLDINBDGLDLAGAIGNAVITLMSRNVQJNSLHPEBPKI	652
QY	660	SVVORDCRRRGQEA	NCTALTCFYQTSRTPGRHQFMYRFTASJDEWTAGARAAPDSG	719
Db	653	NI FHRDCKRSRGA	DCTLAFLCFTPIPLAHPQOTTGVGRYNAVYAMDERRYPRALHDSG	712
QY	720	QRLSPRLRLSVGN	TCQJLHFVILDSYLRPVALTVPALDNTTKGQPVINBESSPSI	779
Db	713	DRFTRAVILSSGQEL	CERINFHVLDTDADYKVPFVSVEYSLBDDH--GPMJLDMPTTL	771
QY	780	QKLVFSDCGPDNC	CVTDIVYQVMMDI-----RQSRK-----APrvv	817
Db	772	RVSVPFMNGCNEDE	HCVPDLVDARSJLPTAMVEQOVRILRKAPODCAVYLSFDTVTI	831
QY	818	RGRGRKVLVSTL	LENRKENAVYTSLSIIFSRMLHSLATTPQRESDIKEYCAAPS--AAH	875
Db	832	ESTRORAVEATLE	NRGENAVSTVINISQASALQFASLILQKEDSOGSIECVNBERLOKQ	891
QY	876	LCSVGHVVFQTGAK	VYTLLEBEFSCSSLLSQVFGKLTASDSLEBNGTLQENTACTSAVI	935
Db	892	VCNVSYPFPRAKA	KAVAFRLDEFKSITFLHLEIELAAGSDNBRDSTKEDVAPLPHL	951
QY	936	QYEPHLLSSSEST	LHRYEHPYGTLP--VGPPEKTKTLRYONUGCYVVSGLIISALLPA	993
Db	952	KYEAADVLTFRSS	SLSHYEVKPNSSJLERYDGIAPPSCTCFIRIONLGLPFIHGMMKITPI	1011
QY	994	VAHGNGYFLTSL	SOVTTN--NASC--IVONLTPGPCPYHBELOHNRJLNGSVTQCOVRC	1050
Db	1012	ATRSNRLILKLRD	FLITDEANISCNWGSTERPRPVE--EDLRARPQLNHSNDVVSJINC	1070
QY	1051	HLGOLAKTEVS	VGLLRVLVHNEFFRRAFKSLTVVSTFELGTBEGSVLQLTASRWSBL	1110
Db	1071	NI-RLVPRQEI	NFHLGLMLRSLKALKYKSMKIMVNALORQHSPIFPEEDPSRQIV	1129
QY	1111	LEVOTRPI	LISMLIGSVLGGILLALVFLCKMGCFPAHKKIPEEBKEBKE	1166

Db 1130 FEISKQEDNQVPIWIIIVGSTIGULLLALVLIALMKIGFFFSAR----RRREFELD 1181

RESULT 9  
AAU14467  
ID AAU14467 standard; Protein; 1188 AA.  
XX AAU14467;  
AC  
AD  
DT 24-OCT-2001 (first entry)  
DE Human novel protein #338.  
XX  
KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;  
KW immunomodulatory; cytosstatic; neuroprotective; vulnerrary; nootropic;  
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
KW tissue regeneration; immune disorder.  
XX  
XX Homo sapiens.  
OS  
PN MO200155437-A2.  
PD  
FD 02-AUG-2001.  
XX  
XX 25-JAN-2001; 2001WO-US02623.  
PF  
PR 25-JAN-2000; 2000US-0491404.  
PX  
PA (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI WPI: 2001-451939/48.  
DR N-PSDB; AAS22772.  
XX  
PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
XX nervous system disorders, and for regenerating bone and cartilage -  
XX  
PS Example 4; Page 828-831; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. Polynucleotides of the invention are used as probes and primers, for sequencing, for chromosome or gene mapping, in the production of recombinant proteins, and in generating anti-sense DNA or RNA and in gene therapy. Polypeptides of the invention can be used to target drugs to a tumour, in assays to determine biological activity, to raise antibodies/elicit an immune response, to determine quantitative protein levels, as tissue markers, and to isolate receptors or ligands. Polypeptides of the invention may also be useful in treating platelet disorders, stem cell disorders, regenerating bone, cartilage, tendon, ligament and/or nerve tissue, wound healing, treating burns, promoting the proliferation, differentiation and survival of stem cells, as a contraceptive, treating osteoporosis and osteoarthritis, anaemia, Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, stroke, immune deficiencies resulting from bacterial, viral or fungal infection or from autoimmunity, cancer, allergy, asthma, graft-versus-host disease, eczema, hemophilia, thrombosis, anti-inflammatroy diseases, nervous system disorders, and infection. The present sequence represents a protein of the invention.

Sequence 1188 AA;

Query Match 39.9%; Score 2438.5; DB 22; Length 1188;



Best Local Similarity 43.2%; Pred. No. 9.2e-208;  
Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;

```

QY 1 MELPFVTHFLPLVFLTGCPSPNDEHHPRLPGGPBEAFGVSYLQHGQGMWLVGA 60
Db 1 MDLPKGLVAVMALSLMPGFTDTFNMTRPRVIPSRTAFPGYVQOQHISGNKMLVGA 60
QY 61 PWDGSGDRGDVYRCFVGAHANAPCAKGLGDYQIGNSHPAVMHLGMSLLETDDGCG 120
Db 61 PLETNGYQKTGVYKCPV---IHGNTKLNLRVTLNSVSEKDMNRLGLSLATPKDMS 117
QY 121 PMACAPLWRACGSSVFSSGICARVDASFQPGSLAPTAQCPTVMVIVLDGNSIYP 180
Db 118 FIACSPFMSHECGSSYYTGMCSRVNSNFRFSKTVAPALQRCQTYMDIYIVLDGNSIYP 177
QY 181 WSEVOTFLRLVGLKFIIDPEQIOVGLVOYGESVHEHMSLGDRTKEVVRRAKNSRRRG 240
Db 178 WVEVQHFLNITLKKFYIGPGQIQVGVQYGEDVHEHFLNDRSVKDVVEAASHLEQRCG 237
QY 241 RETKTAOAIMVACTEGFSQSGHGRPEARLLVVTGDSHDEBELPALKACEAGRVTRY 300
Db 238 TETRTAFGIEFARSEAFQK--GGRKAKKVMIVITDGESHDPDLEKVIQSEBRDVTRY 295
QY 301 GIAVUGHYLRQRDPSFLREIRTTASDDERFENVVTDEALTDIVDALGRITGLGSG 360
Db 296 AAVAVIGYVRRGINPETFLNEIKYIASDDDDKHFFVVTDEALKDIVDALGRITGLEST 355
QY 361 HANENSGFLMSQIGFSTRHLKDGILFGMVAGYVMGSGVLMLEGGHRLFPFRMALDEGF 420
Db 366 N-KNELTSFELMSQIGFSSHVVEDGVLLGAVAGYDMNGVLEKTSAGKIPILRESTLKEF 414
QY 421 PPALONHAAVLGYSSMILRGRRLLFSGAPFRHGRKVIAFOLKDGAVVAOSLQGE 480
Db 415 PEELNKHGAYLTGYTSVSSRQGRVYVAGAPRPHNTGVILFTMHNNRSLTIHQAMRQ 474
QY 481 QIGSTFGEELCPLDTRDGTVDVLLVADPMFLGPQNKETGRVYVYLVGQSLLTQGTIQ 540
Db 475 QIGSYFGESEITVDIDGDTVDVLLVGAPEYFN--EGREGRKYYVEL--RQNRFFVNGTLK 532
QY 541 PPPP--ODARFGFAMGALPDLNODGFADVAVGAPLEGCHGALYLVHGTSQGRPHPAQRI 599
Db 533 DSHSYONARFGSSIASVRDLNODSYVDVAVGAPLEBNHAGIYIFHGFEGSLTKTPKQRI 592
QY 600 AAASMPHALSYGRSGVGRDLDDGDLVAVAGAOGAAILSSRPVHLTPSLLEVTPQRI 659
Db 593 TASELATGQYFGCGSIHQGLDINEGDLIDLAVALGAMNAVILMSRPVQGINASLHPEPSKI 652
QY 660 SVVQDRCRRRGEAVCLTALCFQVTSRTPRGWDHOFYMRFTASLDEWTAGAAAFDQSG 719
Db 653 NIFHDDCKRSGRADATCLAAFLCFTPIFLAPHEQTTVGIRYNAVATMDERYTBRAHLEGG 712
QY 720 QLSLRRLRLSYGVNTCEQMFHVLDTSTYLKPVALTIVFALDNTTKPKPVUNESPSI 779
Db 713 DFTNRRAVLSSGOELCERINFHVLDTADVAVKVPFSVSESLDDPDH--GPMLLDDGAPPTL 771
QY 780 QKLVEFSKXDCGPDNCEVTDLVQLQVMMDI-----RGSRK-----APFV 817
Db 772 RVSVEFGMNCNDEHCVPDLVLDARSDLPTAMEYQCRVLRKAQDCSATLTFDTVTFI 831
QY 818 RGRRRKVLVSTLLENKENAVNTSLISFRMLHLASTLPQRESPIKVECAAP--AAAR 875
Db 832 ESTROVRVAEATLENGENAVSTVINISQSANILOFASLLQKSDDSGSIQVNEERRLQK 891
QY 876 LCSVGHVPOFGAKTFLLEFEFSQSLSLQYFGKLTASDSELENGTQENTAQTSAXI 935
Db 892 VCNVSPFPFRAKAKAFRLDPEFSKSIPLHLEITELAAQSDSNERSTEDVAVAPRFL 951
QY 936 QYEPHLLFSESSTLHREYVAPYGLTP--VGPGPEFTTLRVONTLGVYVSGLILSALLPA 993
Db 952 KTEVDLFLTRSSLSLHYEKPNSSLERVYQIGPFECIRIQLGLGFHGMKMTIFI 1011
QY 994 VAHGNYFLSLSQVITN--NASC-IYQNTTEPPGPVHBEELQHTNRLNGSTQOCQVRC 1050

```

```

Db 1012 ATRSGNRLKLKRLDTDEANTSCNIGWSTYRPTPVE--EDLRAPQLNHSNDVVSINC 1070
QY 1051 HLGLQAKTEVSGVGLRLVHNEFPFRAKFSLTVSTPELGTBEGSVQLTEASRWSSESL 1110
Db 1071 NI-RLVPRQELNFHLGLNLMRLSKALKYKSMKIMVNAALQROFSPPIFEEDSRQIV 1129
QY 1111 LEVQTRPILSLMWILGSLVGLLLALLVFCMLKLGFPFAHKKIPEEKKEKLE 1166
Db 1130 FEISQEDMQVPIMIVIVSTLGLLLALLVIALMKLGFFRSAR---RRREPGLD 1181

RESULT 10
AAB25582
ID AAB25582 standard; Protein; 1189 AA.
XX
AC AAB25582;
XX
DT 21-NOV-2000 (first entry)
XX
DE ITGA11 protein encoded by human secreted protein gene #7.
XX
KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
KW antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
KW anticancer; vulnery; antiviral; antibacterial; antifungal;
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW Crohn's disease; nephritis; hyperproliferative disorder;
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
XX
OS Homo sapiens.
XX
WO200029435-A1.
XX
PD 25-MAY-2000.
XX
PF 27-OCT-1999; 99WO-US25031.
XX
PR 28-OCT-1998; 98US-0105971.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y,
PI Greene JM,
XX
DR MPI: 2000-387742/33.
XX
DR N-PSDB; AAB80612.
XX
PT Isolated nucleic acid molecules encoding human secreted proteins are
PT used for the prevention, amelioration and treatment of autoimmune,
PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
PT wounds, and infectious diseases -
XX
PS Claim 1; Figure 19A-F; 803pp; English.
XX
XX

```

The present invention relates to 12 secreted human proteins and the nucleotide sequences encoding them. The polynucleotide sequences given in AAB80606-180623 encode the 12 secreted protein sequences given in AAB25576-825593. The human secreted proteins have various activities dependent on the tissues in which they are expressed. Examples of the activities of the proteins include: immunosuppressant; anti-inflammatory; antiarthritic; antirheumatic; dermatological; antiproliferative; antiarteriosclerotic; anticancer; vulnery; antiviral; antibacterial; and antifungal activity. The proteins, polypeptides, agonists and antagonists may be used to treat prevent and/or diagnose various diseases, disorders and conditions examples of which include: immune disorders e.g. Addison's disease, rheumatoid arthritis, dermatitis, and multiple sclerosis; inflammatory disorders e.g. inflammatory bowel disease, Crohn's disease and nephritis; hyperproliferative disorders such as paraproteinemias and purpura; cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis; cancer e.g. melanoma and lymphoma. The proteins and polynucleotide sequences may also be used in wound healing and the treatment of

CC Infectious diseases. The human secreted protein gene #7 and protein  
 CC sequences are represented in sequences AAA80612 and AAA25582. Secreted  
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences  
 CC AAA80652-AB80661 represent genes which are related to the secreted protein  
 CC gene#7.  
 CC  
 XX  
 SQ Sequence 1189 AA:  
 Query Match 39.8%; Score 2429; DB 21; Length 1189;  
 Best Local Similarity 43.1%; Pred. No. 6.5e-207;  
 Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;  
 QY 1 MELPFTHLFLPLVLTGLCSPEFLNDHHRLLPFGPEAEFGYSLQHWGGGQRMVLVGA 60  
 DB 1 MDLPRLVAVAMALSLWPGFTDFTNMDTRKPRVIGSRTAFYGTVOQHDSIGKMLVGA 60  
 QY 61 PMDGSGDRGDYRCPVGGANAPCAKGLGDYQIGNSHPAVNNHLGMSLLETGDDG 120  
 DB 61 PLETNGYOKTGDYVKKCPV--IHGCTKMLGRVTLISNVSERKDNRLGLSLATNPKDMS 117  
 QY 121 FMACAPLMSRACGSSVFSGICARVDASFOQSLAPTQRCPTVMDVILDGNSIYP 180  
 DB 118 FLACSLMHSHEGSSYTTGMCGRVSNFRFKTVAPALORCQTYMDIYIVLDGNSIYP 177  
 QY 181 WSEVQTLRLVYKLFIDEQIQVLVQGESPVHMSLDFTKEEVYRAAKNLAREG 240  
 DB 178 WVEVQHFLINILKKFYIGQIQVGVQGEDVHVEFLINDYASVDVVEASHIEQGG 237  
 QY 241 RETKTQALMVACTEGFSQSHGGRPEARLLVYVTGESHGDELPAALKACAGAVTRY 300  
 DB 238 TETRTAFGIEFARSEAFQK--GGRKGAKKVMYITGESHSDSDLEKRVQOQSRDVTY 295  
 QY 301 GIAVLGHYLRQRPDPSFLREIRTIASDPDERFFVNTDEAALTDIVDALGRIFGLGSS 360  
 DB 296 AVAVLGYVNRGNGPFLNEIKYIASDPDDKAFVNTDEAALKDVIDALGRIFGLGSS 355  
 QY 361 HAENESSFGLMSQISFSTRKLDKGLFCGVAGYDWGCVLMEGRLLFPPRMALDEF 420  
 DB 356 N-KNETSFGLMSQTFSSHVEDVLLGAVGADWNGAVLKETSAGKYPILAESYLKEF 414  
 QY 421 PPLQNHAAVLGYSSSMLRGSRLLFLSGAPFRFRKGVIAQLKKDQAVVAQSLQGE 480  
 DB 415 PELLKNHGAVLGYTVSVSSRGRRVYVAGAPRFNHTGKILFTMHNNSSLTTHQAMRQ 474  
 QY 481 QIGSYGSELCPTDTRDGTDTDLVLAAPWFLGPKMKETGRVYVYVVGQSSLTTLTGTO 540  
 DB 475 QIGSYGSELTVDIDGDTVDVLLVGAIFYFN-EGREKGVYVEL-RQNRVYNGITLK 532  
 QY 541 PEPP-QDARFGFAMGALPDLNODGFADVAVGAPLEDDHOGALVLYHGTSQVAPHPAQRI 599  
 DB 533 DSHSYQNAFFGSSIASVRDLNODSYNDVVYVGAFLLEDNHAGAIYIFHGFSGSLIKTEKORI 592  
 QY 600 AAASMHALSYFRSVDGRLLDGDLDVVAVGAQAAIILSRPIVHLTPSLSEVTPQAI 659  
 DB 593 TASELATGQYFCGSIHQGLDLMEDGLIDLAVAGLANAVLMSRPVQVQINASHPEPSKI 652  
 QY 660 SVVQRCRRGCGAVCTAALCFQVTSRTPRGWDHGFYVRFSTLSEMTWAGAAAPADGSG 719  
 DB 653 NIFHRCKCKSGRATLAAFLCTPTPLFAPHFQTTVGIRYKNTMERKYPRAHLDEGG 712  
 QY 720 QRLSPRLRLSVGNVTCEQLHFVLDTSVLRVALVTYFALDNTTKPGVNLGSGPTSI 779  
 DB 713 DRFTNAVALLSGQELCERINFHVLDTADVKKVFTSVESYLEDPDH-GPMLDDGPTTL 771  
 QY 780 QKLVPSKCKGPNNECVTLVLOVMDI-----RGRK-----APPV 817  
 DB 772 RVSVPFWNGCNEDEHCVPLVLDARSDPLTAMEYCORVLRKPAQDCSAYTLSEDTTVFI 831  
 QY 818 RGRGRVAVLSTDENKENAVNTSLITFSRNLHLASLMPQSPKIVCAAPS-AAHR 875  
 DB 832 ESTROAVAVATLENGENNYSIVLWISQANIQFSLQKEDSDSIECVNEERLQKQ 891  
 QY 876 LCSVGHVFPQTGAQVTFLEFEFSCSLISQVFGKLTASSDSLERNGLTQENTQTSAYI 935

DB 892 VCNVSYPPFRKAKAVAFRLDPEFSKSIPLHHLIELAAGSDSNBDSKEDVAVBLRPHL 951  
 QY 936 QYEHLLSSSESTLHRYVHPYGTLP--VGPEPEKTLIRVONIGCVYSGIILALLPA 993  
 DB 952 KYEADVLFTRSSSLSHYEKLNSSLERYDGIQPPSCIFRIONLPLPHIGIMKITPTI 1011  
 QY 994 VAHGNVFLSLSOVNTN--NASG-IYONLTPEPPVAPPELQHTNRNGSNTOCQVVR 1049  
 DB 1012 ATRSGNRLKLKRLDPLTDEVANTSCNMGNSITFRTPVE-EDLRAPQLNHSNDSVGIN 1070  
 QY 1050 CHLGQLANGTEVSVGLRLVHNEFFRAKPKSLTVVSTFELGTEEGSVQLTTEASRWSBS 1109  
 DB 1071 CNI-RLVNPQEIHFLLGNLMLRSIKAKYKSMKIMVAAALQRPCHSEFIFREEDPSRQI 1129  
 QY 1110 LLEVQTRPILSLWLLGSLVGLLALLLVFCIMKGFFAHKKIPEEKREKLE 1166  
 DB 1130 VFEISKQEDWQVPIWIVGSTIGLLALLVALLMLKLGFFRSAR---RRREPGLD 1182  
 RESULT 11  
 ABG12949  
 ID ABG12949 standard; Protein; 1189 AA.  
 XX ABG12949;  
 AC  
 XX  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #12940.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Dmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AA571136.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 43308; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1189 AA:

Query Match 39.8%; Score 2429; DB 22; Length 1189;  
 Best Local Similarity 43.1%; Pred. No. 6.5e-207;  
 Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

QY 1 MELPVTHFLPLVLTGLCSPPNIDHRLPPGPPEAFGYSYLQHGCGQRMVYGA 60  
 DB 1 MDLPGLVYVAMALSLMPGFTDFNMMDTRKPRVYPSGRTAFFGYTQOHDISGNKMLVYGA 60  
 QY 61 PNDGSGDRGDRGVRRPVGGAHNAPCAKHLDYQLGNSHRAVNMHLGMSLLEDDGG 120  
 DB 61 PLETNGYOKTGVYKCPV---IHGNTKLTNLGRVTLISNYSERKDNMRGLSLATNPKNDS 117  
 QY 121 FMACAPLMSRACSSVFSVSGICARVDASFOPOGSLAFTAQRCPTVMDVITVLDGNSIYP 180  
 DB 118 FLACSPBMSHECGSSYTTGMCSSRVNSNFRPSKTVAPALQRCCTYMDIYIVLDGNSIYP 177  
 QY 181 WSEVOTFLRLVQKLFIDPEQIQVGLVYGESPVHEMSLGDPRTEEVRRAKNLSRREG 240  
 DB 178 WVEVQHFLINILKFIYIGQIQVGVVQGEVDVHEFHLNDRSVADVVEASHIEQRG 237  
 QY 241 RETKTAQAIMVACTBEGFSQSHGPREAALLVVTVDGESHDBELPALKACEAGRVTRY 300  
 DB 238 TETRTAFGLEFARSEAFQK--GGRKGAKKVMIVITDGBSHDSDLEKVIQSERDNTVY 295  
 QY 301 GIAYIGHYLRORDPSSFLREIRITIASDDEPFPFVNTDEALVTIDVADGRIFGLEGS 360  
 DB 296 AVAVLGTYNRKGINPETFLEIKYIISDDDKFFNVTDKALKVIDLADGRIFSLBET 355  
 QY 361 HAENESSFGLMSQIGFSTHRLKDGLFGWGAAYDWGSSVLWEGHRLFPPRMALDEDF 420  
 DB 356 N-KNETSFGLMSQIGFSSHVVEDVGLGAVGYDMNGAVLKETSAGKVIPLRESYLKEF 414  
 QY 421 PALQNHAAVLYGVSYSMILRGRLFLSGARFRFRKGVIAFOLKKQDAVVAQSLQGE 480  
 DB 415 PEELKNHAGLYTYIVSYVSSRQGRVYAGAPRFNHTGVILFTMNNNSLTTHQAMRQ 474  
 QY 481 QIGSYFGEELCPIDRDGTTDVLVAAPMPLGPQNKERTGRVYVYLVGQSSLLTQGTQ 540  
 DB 475 QIGSYFGEELTSDVDGDTVDVLVGAEMVTRN-BERERKQVYIEL-RQNRFTYNGILK 532  
 QY 541 PEPP-QDAFFGAMGALPDINODGFADVAVGAPLEBDHGALYLYHGTSQGVRRPAPRI 599  
 DB 533 DSHSYQNAFSSIASVRLNODSYNDVVGVAFLBNHAGALYIPIHGFPGSLIKTPKQRI 592  
 QY 600 AAASPHALSYGRSVYDGLDLDGDDLVVAVGAGACALILSSRPVHLTSPLEVTPOAI 659  
 DB 593 TASELATIGQYFGCSIHQDLNEDGLIDLAGALGNVAILMSRPVQINASLHFEPSKI 652  
 QY 660 SVVQDDCRRRGOEAVLTALCEQVTSRTPGMDHOFVWRFTASLDEWTAGARAFDSSG 719  
 DB 653 NLFHEDCKRSGRDATLALFLCFTPIFLAPHOTTTVGIRYNATWDERRYTPRAHIDEG 712  
 QY 720 QRLSPRLRLSVGNVTCQQLHFHVLDTSDYLRFVALTVFALDNTTKPGPVLNCSPTSI 779  
 DB 713 DFTNRALVILSSGQELCERINHFHVLDTAYVKVFESVYSLEDDPH-GPMLDDGMPTLL 771  
 QY 780 QKLVPPSKCGPNDNCTDLVLTQVMMDT-----RGRK-----APRV 817  
 DB 772 RVSVPPMNCNDEDEHCVPLVLVARSDDLPTAMEYCORVLRKPAODCSAYTLTSPDTVFII 831  
 QY 818 RGRARRVLSTILENKENAVNTSLSIIFSRNLHLASLPORSPKIVECAPS-AAAR 875  
 DB 832 ESTRQAVAVATLENNGENAYSTVLTNISQANLQFASLIQKEDSDGSLTCVNEERRLOK 891

QY 876 LCSVGHVPGVQCAKTFLLERFSCSSLLSOVFGKLTSSSLSBERNGTLQENTQATSAYI 935  
 DB 882 VCNVSYPPEFRAYKAKAFRLDEPFSSKIFLHLEIEIAGSDSNBERSTKEDNVAPLRPHL 951  
 QY 936 QYEPHLFSSSSTLHRYVHVHTLP--VGPGEKTKTLRQNGCYVAVSGIISALLPA 993  
 DB 952 KTEADVLTFRSSSSSHYEVKKNLSLERYDGPPECTFRIQNLGLFPHGMMKTTITFI 1011  
 QY 994 VAHGNVFLSLSOVITN--NASC-IVQNLTEPPGPVPHPELOHTNRLANGSNTQOQVVR 1049  
 DB 1012 ATRSGNRLKLKRDFLTDEVANTSCNMGNSSTREYRTPVE-EDLRAAPQLNHNSNVASIN 1070  
 QY 1050 CHLGQLAGTGVSGCLRLVHNEFFRAKFKSLTVSTFELGTBEGSVLQLTASRMSBS 1109  
 DB 1071 CNI-RLVNOEINFFLGNLWLSLKALKYKSMKIMVAAALQROFHSPIFREEDPSRQI 1129  
 QY 1110 LLEVQTPILISMLILGSLVGLLALLVFCMLKGFPAHKKIPEEKREKLE 1166  
 DB 1130 VFEISKQEDQVPIWITVIGSTLGLLALLVLALWKLGFRRSAR---RRREPGLD 1182

# RESULT 12

AAB50087  
 ID AAB50087 standard; Protein, 1188 AA.

XX AAB50087;

DT 19-MAR-2001 (first entry)

XX Murine A259.

KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;  
 KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;  
 KW rheumatoid arthritis.

OS Mus sp.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Location/Qualifiers  
 1..22  
 /label= Signal\_peptide  
 23..1188  
 /label= Mature\_protein  
 1..1141  
 /label= Extracellular\_domain  
 39..74  
 /label= Integrin\_alpha\_subunit\_repeat\_domain\_#1  
 115..157  
 /label= Integrin\_alpha\_subunit\_repeat\_domain\_#2  
 164..345  
 /label= Integrin\_alpha\_subunit\_repeat\_domain\_#3  
 367..392  
 /label= Integrin\_alpha\_subunit\_repeat\_domain\_#4  
 421..455  
 /label= Integrin\_alpha\_subunit\_repeat\_domain\_#5  
 478..516  
 /label= Integrin\_alpha\_subunit\_repeat\_domain\_#6  
 540..575  
 /label= Integrin\_alpha\_subunit\_repeat\_domain\_#7  
 602..640  
 /label= Integrin\_alpha\_subunit\_repeat\_domain\_#8  
 1142..1164  
 /label= Transmembrane\_domain  
 1165..1188  
 /label= Cytoplasmic\_domain

WO200073339-A1.

07-DEC-2000.

15-MAY-2000; 2000WO-US13262.

28-MAY-1999; 99US-0322790.

27-APR-2000; 2000US-0561263.

XX (MILL-) MILLENNIUM PHARM INC.  
 PA Pan Y, Lora JM;  
 PI WPI, 2001-041142/05.  
 DR N-PSDB; AAC91904, AAC91905.  
 XX  
 PT Nucleic acid encoding alpha-integrin subunits, useful for treatment and  
 diagnosis of fibrosis, e.g. of the liver  
 PS Claim 8; Fig 5; 164pp; English.  
 XX  
 CC The present sequence is murine integrin alpha subunit, A259 is  
 homologous with the alpha1 and alpha10 integrin subunits and is  
 overexpressed in fibrosis. A259 is implicated in regulation of  
 proliferation, differentiation and/or function of many different cell  
 types. Inhibitors of A259 activity are useful for the treatment of liver  
 disease, particularly fibrosis, and also fibrosis in other organs  
 CC (specifically lung and kidney). In addition, A259 can be used for  
 treatment and prevention of cancer, osteoporosis, acute myeloid  
 CC leukemia, HIV infection, and rheumatoid arthritis.

XX Sequence 1188 AA;

Query Match 39.7%; Score 2422; DB 22; Length 1188;  
 Best Local Similarity 43.4%; Pred. No. 2,7e-206;  
 Matches 513; Conservative 207; Mismatches 412; Indels 50; Gaps 15;

18 GLCSPPNLDHHRFLFPGEPEAFEGYSVLQHVGGGRMLVGAPEWDGSGDRRGDYRCP 77  
 18 GFDFDTNMDTRNRYAGSPSAAFEGYTVQOHDSIGKKMLVGAPEMTNGHOKTGVDYKCP 77  
 78 VGGANAPCAKGLGDIYQVGNSSHPAVNMHLGMSLLETDDGGGFMCAPLMSACGSSVF 137  
 78 VTQGN--CTKMLNGVTLTNSVSEKDNRLGLSLTNPKNDSFLCSPLMSIECCSSYX 134  
 138 SSGICARVDASFPQGSILPTAQRCPTVMVDVIVLDGNSIYVSEVOTFLRLVGLFI 197  
 135 TTMGMSRVNSNFRFSKTVAPALORCOTYMDIVIVLDGNSIYVSEVOTFLRLVGLFI 194  
 198 DEBOIQVLVOYGESVPHWESLGDFTKEEVRAAKNLSRREGRETKTAQAIWACTEGF 257  
 195 GEGQIQVGIYQVGEDVAFHFLNDYRSVDVVEAASHIQRGGTETRTAFGLEFARSEAF 254  
 258 SQSHGGRPEAARLLVVVTGESHDSDELPAALACAGVTRGIVLGHVLRQDPSPS 317  
 255 QR--GGRKAKKVMIVITDGESHDSPLLEKVRQSEKDNVTRAVAVLGYNRGINPET 312  
 318 FLREITIASDPERFFNVTEAALTDIVDALGDRIFGLEGSHAENESSFGLGMSQIGF 377  
 313 FLNEIKYIASDPDKKFFNVTEAALKDIDVADGDIIFLEEGN-KNEISFGLGMSQIGF 371  
 378 STHRLKDGILFGMVGAYDMGVSVMLEGGHRLPPPMALDEDFPPALQNHAAVLGYVSS 437  
 372,SSHVEDGILLGAVGAYDMNGAVLKETSAGKVIPIHRESYLEKEPEBELKXNAAYLGTYVS 431  
 438 MLRGGRRLFLSGAPRRHKGVIARQKKDGAVRVAQSLQSGQISYSGSELCPIDTOR 497  
 432 VVSSRGRRVYVAGAPRRHGTGKILFSMHNRRSLTHQALRGQISYSGSEITSDVAND 491  
 498 DGTVDVILVAAPMFLPQNKETGRVVYVLVGQSSLLTLOGLPBBP-ODARFGFMGL 556  
 492 DRTVDVILVAGAPMYS-BGRERKVVYVNL-RNRNVVYGTGLKDSHSYONARFGSIAV 549  
 557 PDINODGPDVAVAGAPLEDGHCAGALYLYNGTOSGVRRPRAQRTAAASMPHALSYGRVD 616  
 550 ODINODSYNDVVVAGAPLEDHSHRGAIFYFGFQTNILKKPMORITABELAPGLOHFCSTH 609  
 617 GRDLUDGDLVAVAGAGAAIILSSRPVHLPSLEVPQALISVQDRRCRGCAVLT 676  
 610 GOLDLNEGDVLVAGALGNAAVLMARPVQVIMASLHFEPSKINIPHKDCKRGRDATCL 669

QY 677 TAAICFOVTSRTPGEMDHOQWIRFTASLDEWITAGARAAFDGSGORLSRRLSVGNVTC 736  
 DB 670 AAFICFIFILAPHPQATVAGIRVATWDERRYMRRAHLDEGGDQFTNRRAVLLSSGQHC 729  
 QY 737 EQLHFRVVDTSYLRPAVALLTFPAIDNTTKGQPVINESSPSTISQVLVPEFSKDCGDNICV 796  
 DB 730 QRINHVVDIADYVPAVSVSEYSLBDPN-GPMLDNGMPTTLRVSVFPMNGCNEDHCV 788  
 QY 797 TDLVQNMVDIRGSK-----APFVVRGRKRVLTSTLENRK 834  
 DB 789 PDLVDASDLPFTAMEYQORVLGRAPQDCSSYTLSTFDTTVPIESTRRVAVALELRNG 848  
 QY 835 ENAVNTSLSIIFSRNLHSLTPRESPIKYECAPS--AHARLCSVGPVFOGAKYTF 892  
 DB 849 ENAVAVANTISQSENLQFASLIQKDSDNSIECVNEERRLHKVGNVSYPPRAKAKYAF 908  
 QY 893 LIEFFSCSSLSQVFGKLTASSDSLEKNGTQENTAQTSAYIQVEPHLLSSSESTLRY 952  
 DB 909 RLDFFESKSVPLHQLQIHGAGSDSHEDSTADDTALLRPLKYEADVLFPRSSLSHF 968  
 QY 953 EVHPYGTLPV--GPGEPEKTLRYONLGCYVVGIIISALLPAVHAGNYFLTSQVITN 1010  
 DB 969 EVKANSLSLESDIGIEPPNCCYFKVONLGFPIHGVMMKITVPDIATRGGRNLLMDPFTD 1028  
 QY 1011 --NASC-IVONLTEBPGRPVHPEELQHTNRNGSNTOQOVVRCHLGOLAKTEVSGLLR 1067  
 DB 1029 QCNSTCNWGNSTERYSTPE-EDLSHAPQRNHSNDSVSIICNL-RLAPSQETSFYLVG 1086  
 QY 1068 LVHNEFFPRAKFKSLTVSTFELGTEBGSVLQLTASWSSLSLEVQTRPILISLMLI 1127  
 DB 1087 NMLTSLALKYRSLSKIVNALQRFHSPFIREDSDRQVTEIISKQEDMQVPIWITV 1146  
 QY 1128 GSVLGLLLALLVFCMLKLGFAHK-----IPEEEK 1160  
 DB 1147 GSTLGLLLALLVLAHMKLGFFKSAKRKRREGLCPTKEIK 1188

# RESULT 13

AAU10552  
 ID AAU10552 standard; Protein; 1188 AA.

XX AAU10552;

XX 14-FEB-2002 (first entry)

XX Murine A259 polypeptide.

XX Mouse, A259; integrin alpha subunit; integrin alpha 10; secreted protein;  
 KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;  
 KW cartilage associated disorder; haematopoietic disorder; bone marrow;  
 KW immune related disease; apoptotic disorder; neuronal tissue disease;  
 KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;  
 KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;  
 KW antidiabetic; antihemodialytic; antiallergic; antisthmatic; dermatological;  
 KW antidiabetic; anticonvulsant; antiparkinsonian.

OS Mus musculus.

XX Key

XX Peptide

XX Domain

XX Protein

XX Domain

XX Domain

XX Domain

XX Domain

Location/Qualifiers  
 /note= "Signal peptide"  
 1..1141  
 /note= "Extracellular domain"  
 23..1188  
 /note= "Mature murine A259"  
 39..74  
 /note= "Integrin alpha repeat domain"  
 115..157  
 /note= "Integrin alpha repeat domain"  
 164..345  
 /note= "I domain or Von Willebrand Factor type A domain"  
 367..392  
 /note= "Integrin alpha repeat domain"

```

FT  Domain 421..455
FT  /note= "Integrin alpha repeat domain"
FT  Domain 478..516
FT  /note= "Integrin alpha repeat domain"
FT  Domain 540..575
FT  /note= "Integrin alpha repeat domain"
FT  Domain 602..640
FT  /note= "Integrin alpha repeat domain"
FT  Domain 1142..1164
FT  /note= "Transmembrane domain"
FT  Domain 1165..1188
FT  /note= "Cytoplasmic domain"
FT  WO200181414-A2.
FT  01-NOV-2001.
FT  27-APR-2001; 2001WO-US1516.
FT  27-APR-2000; 2000US-0561263.
FT  (MILL-) MILLENNIUM PHARM INC.
FT  Pan Y, Lora J;
FT  WPI; 2002-041397/05.
FT  N-PSDB; AAS16874.
FT  New A259 nucleic acids and polypeptides, which comprise integrin alpha
FT  subunit, useful for diagnosing, preventing or treating e.g. liver
FT  disease, kidney or lung fibrosis, cancers, blood disorders or immune
FT  related diseases
FT  Claim 9; Fig 5; 168pp; English.
XX  The invention relates to human and murine A259 nucleic acid molecules
XX  which encode secreted proteins with homology to integrin alpha subunits,
XX  specifically to integrin alpha 10. The A259 polypeptide and nucleic acid
XX  are useful for treating liver disease or fibrosis, particularly kidney
XX  fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also
XX  useful for diagnosing, preventing or treating cartilage and bone
XX  associated disorders (such as bone cancer, achondroplasia, myeloma,
XX  fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and
XX  osteoporosis), bone marrow, blood and haematopoietic disorders (such as
XX  acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune
XX  related diseases (such as HIV viral infections, cancers, T cell
XX  autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.
XX  asthma and psoriasis), apoptotic disorders (such as systemic lupus
XX  erythematosus and insulin-dependent diabetes mellitus), diseases of the
XX  neuronal tissues (such as epilepsy and muscular dystrophy) and
XX  neurodegenerative diseases (such as Parkinson's disease and Huntington's
XX  disease). This sequence represents the murine A259 polypeptide.
SQ  Sequence 1188 AA;
Query Match 39.7%; Score 2422; DB 23; Length 1188;
Best Local Similarity 43.4%; Pred. No. 2, 7e-206;
Matches 513; Conservative 207; Mismatches 412; Indels 50; Gaps 15;
QY 18 GLCSFNDLHHPLFPPEPEAFSGSYLVQHVGGQRMWLVGAPMDGSGDRGDDVYRCP 77
DB 18 GFTDFNDMDTRPRVIAQSAFAFGVTVQOHDISGKWLVGAPMETGHCQKTGDVYKCP 77
QY 78 VGGAHNAACAKGHLDYDQNGSHPAVMHILGMSLFTDDGGFACAPLWMBAGSSVF 137
DB 78 VTQGN---CTKLNLRVTLSNVSEKDNRLDLSTATPKDPSFLACSPWMSHECGSSSY 134
QY 138 SSGICARYDASFQPOGSLAPTAQRCPYVDVIVLDGNSIYPMSEVOTFLRLVGLKFI 197
DB 135 TTGMSGRVSNFRFKTAPALQRCQTYMDIVIVLDGNSIYPMVEVGHFLNLIKKFI 194
QY 198 DPEQIQVGLVQYGESPVHEMSLGFRTKEEVVRAAKNLSREGRTKTDQAIMVACTEGF 257

```

```

DB 195 GPQGIQVGIYGEDVAHEFLNDIRSVKDVVEASHIEQRGNETRTAFIEFARSEAF 254
QY 258 SQSHGRPREARLLVVTVDGESHDEBELPALKACAGRVTVYGIAGLHARLRORDPS 317
DB 255 OK--GGRKGAKKWMLVITDGESHDSPLDEKXIRQSEKDNVRYAVAVLYGNKGINET 312
QY 318 FLREIRTIASDPDERFFPVNTDEALTDIDALDRIFGLEGSHAENESSFGLMSQIGF 377
DB 313 FLNEIKYIASDPDDKHFFNVTVDEALAKDIDALDRIFSLBGTN-KNETSFGLMSQIGF 371
QY 378 STHLKQGIILFGMGANDYMGSSVYMLBEGHILFFPRMALBDEFPFALONHAAYIGYSVS 437
DB 372 SSHVEDDILLGAAGVDMNGAVLKETISAGVYIPHRESYLKEPPEELKNHAAYIGYVTS 431
QY 438 MLNRGRRLFLSGAPRFRHGRKVIAFOLKKGAVRVAQSLQGEQISYFGSELCPLODR 497
DB 432 VVSSRQGRVYVYAGAPRFRHGTGKVLFSMHNKRLTIHOLAGQDQISYFGSEITISVDND 491
QY 498 DGTIDVLLVAAPELGPONKETGRVYVYLVGQOSLTLTQTLQDEPP--QDARFGFAMGAL 556
DB 492 DRVTIDVLLVAPWYFS--EGREGRKVVYVNL--RQNRFYVNGTLKDSHSYQNAFSGCIASV 549
QY 557 PDLNODGFADYAVGAPLEDGHQGLYLYHGQSGVRPHPAQRIAAASPHALSYGRSVD 616
DB 550 QDLNODSYNDVAVGAPLEDGSHRGAIFYFHGFOTWLKPKMORITASELAPGLQHFQGSIH 609
QY 617 GRLLDGDPLDVAVAGQGAAILLSRPVYLTSPLEVTPOAISVVOGRDRCRGOEAVCL 676
DB 610 GQDLNDEGLVDLAVAGLGNVAVLMARVYVINSLHEPESKINIFHKDCRNGRDATCL 669
QY 677 TAAICFOVTSRTPGRMWHQFYMRFTASLDEWTAGARAFDGGGQRLSPRLRLSVGNVTC 736
DB 670 AALCFEPIFLAPHFQATVGIIRYNATWDERVMPRAHLDGQGFTRAVALLSSGQDHC 729
QY 737 EQLHFRVLTDSYLRPAVALTYTPALDNTKGPVNLBESPTISQKLVFSDCCGDNCEV 796
DB 730 QRINFHVLTDDYVKAFAVFSVEGLBEDPDN--GPMLDNGMPTLTVSVSPFMNGCNEDEHCV 788
QY 797 TDVLVQVMDLRGSRK-----APVNVGGRKVLVSTLERK 834
DB 789 PDLVLDARSDELPTMEVYQVRVLRPADCCSYTISFTTTFIISTRRRVAVETLENRG 848
QY 835 ENAVNTSLIIFSNHMLASLTPQRESPIKVECAAPS--AAHRLCSYGVHPVQTCARVTF 892
DB 849 ENAVSAVLNIGQSNLQFASLIQDSDNSIECVNEERLHKYCNVSPFFRAKAKYAF 908
QY 893 LLEFPGSCSLLSQVFGKLTASSPSLERNGLQENTQTSAYIOTYEPHLLFSSSTLHRY 952
DB 909 RLDEFKSVPLHNLQIHLGAGSDSHRODSTADNTALRLRHLKYEADVLTFRSSSLSHF 968
QY 953 EVHNYGTLFV--GPGPEKKTLYRQNLGCVVSGLLIISALLPAVAHGNGVFLISQVITN 1010
DB 969 EVKANSSLESYDGGPPNVCVFKVQNLGFFPIHGVMMKITVPIATRGGRNRLMLRDFFTD 1028
QY 1011 --NASC-IVQNLTEPPGPVYHPELQHTNRLNGSNTQOVVRCHIGQLAKTEVSGILR 1067
DB 1029 QGNNSCINWGSTYRSTFTE--EDLSHAPKNHNSNDVSIICNL--RLAPQDEFISFYLVG 1086
QY 1068 LVHNEFFRRRAKFKSLTVVSTFELGTERGSVQLTQASRWSSSLLEVQTRPILISLWLI 1127
DB 1087 NLWLTSLKALKYRSLIKTVNAAALORPHSPPIFREEBPSPROVTFEISKOEDMOVPIWII 1146
QY 1128 GSVYGLLLALLALVFCMLKLGFFAHK-----LPEEEK 1160
DB 1147 GSTIGGLLLALLVLAALKLGFFKSAKRKEBPGLPIPEKIL 1188

```

RESULT 14  
 AAB30929  
 ID AAB30929 standard; Protein; 1188 AA.

XX  
 AC AAB30929;  
 XX



RESULT 15  
 AAB25590  
 ID AAB25590 standard; Protein; 1034 AA.  
 XX  
 AC AAB25590;  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX  
 DE Protein encoded by human secreted protein gene #7 clone HOBY69.  
 XX  
 KM Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;  
 XX anti-rheumatic; dermatological; antiproliferative; antiarteriosclerotic;  
 KM anticancer; vulnery; antiviral; antibacterial; antifungal;  
 KM immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;  
 KM multiple sclerosis; inflammatory disorder; inflammatory bowel disease;  
 KM Crohn's disease; nephritis; hyperproliferative disorder;  
 KM cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;  
 KM melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.  
 OS Homo sapiens.  
 XX  
 PN MO200029435-A1.  
 XX  
 PD 25-MAY-2000.  
 XX  
 PF 27-OCT-1999; 99WO-US25031.  
 XX  
 PR 28-OCT-1998; 98US-0105971.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JT, Moore PA, Wei Y;  
 PI Greene JM;  
 DR WPI; 2000-387742/33.  
 XX  
 PT Isolated nucleic acid molecules encoding human secreted proteins are  
 PT used for the prevention, amelioration and treatment of autoimmune,  
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,  
 PT wounds, and infectious diseases -  
 PT  
 PS Claim 1; Page 678-682; 803pp; English.  
 XX  
 CC The present invention relates to 12 secreted human proteins and the  
 CC nucleotide sequences encoding them. The polynucleotide sequences given  
 CC in AAB25590-AAB25593 encode the 12 secreted protein sequences given in  
 CC AAB25576-A25593. The human secreted proteins have various activities  
 CC dependent on the tissues in which they are expressed. Examples of the  
 CC activities of the proteins include: immunosuppressant;  
 CC anti-inflammatory; antiarthritic; anti-rheumatic; dermatological;  
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnery;  
 CC antiviral; antibacterial; and antifungal activity. The proteins,  
 CC polypeptides, agonists and antagonists may be used to treat prevent  
 CC and/or diagnose various disease, disorders and conditions examples of  
 CC which include: immune disorders e.g. Addison's disease, rheumatoid  
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders  
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;  
 CC hyperproliferative disorders such as paraproteinemia and purpura;  
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;  
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide  
 CC sequences may also be used in wound healing and the treatment of  
 CC infectious diseases. The human secreted protein gene #7 and protein  
 CC sequences are represented in sequences AAB25592 and AAB25582. Secreted  
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences  
 CC AAB25582-A25593 represent genes which are related to the secreted protein  
 CC gene#7.  
 XX  
 SQ Sequence 1034 AA;

Query Match 36.9%; Score 2254; DB 21; Length 1034;  
 Best Local Similarity 44.5%; Pred. No. 2.2e-191;

Matches 464; Conservative 187; Mismatches 356; Indels 36; Gaps 11;

QY	1	MELPFVTHLFLPLVLTGLCSFNFNDENHRLPFPAPPAPFGSYVLQHVGGGQRMVLVGA	60
DB	1	MDLPRLGVVAAALSLMPGFTDTFNMDTRKPRVIPSRSRAFGCTYQGDHISNKKLVVGA	60
QY	61	PWDGSGRRGDVYRCVPGAHNAPCAKGLGDYQLGNSSHPAVMHLSLLETDGGG	120
DB	61	PLETNGYKGTGDVYKCPV---IHGNCIKLNGRVLTLSVNSERKDMRLGLSLATPDKNS	117
QY	121	FMACAPLMSRACGSSVFSSGICARVDAFQOGSLAPAQRCPTMDVIVYDGSNSTYP	180
DB	118	FLACSPLMSSHGSSVYTTGMCSSRVNSFRSKTAPALQRCQYMDIVIVDGSNSTYP	177
QY	181	WSEVTFPRLVGLKFIPEOIVGLVQYGSVPVEMSLGPTREEVVRAKNSRREG	240
DB	178	WEVEHFLINLIKRYIGRQIQVGVVQYGEVYVHRLNDRSKVDVEASHLEQNG	237
QY	241	RETQTAAIMVACTEGFSQSHGPREARLLVVTVDGESHGDEELPAALKACEAGVTRY	300
DB	238	TETRTAFGIEFARSAFQK--GGRKAKKVMIVITDGESHSDSPDLEKVIQGERDNTRY	295
QY	301	GIANVGHYLRQRDSSFLREIRTAASPDDEFFNNVDEALTDIVDALGRIGLGS	360
DB	296	AAVALGYNNRGINETFLNEIKYIASDDDKGHFNNVTDEALKDVIDALGRIFSLRG	355
QY	361	HAENESSFGLSEMSQIGFSTHRLKDGLFGMVGAYMGGSVLMLEGHRLFPPRMALDEF	420
DB	356	N-KNETSFGLSEMSQIGFSSHVYEDGVLGAVGAYMNGAVLKETISAGKVIPLRESYLKEF	414
QY	421	PPALONMAAYLGYSVSMILRGRRLLFGARFRHRKQVIAFOLKKDGAFFVAASLOGE	480
DB	415	PEELKNHGAIVGYTVTSVSSRQGRVYVAGAPRFNHTQKVLFTMHNNRSLTIHQAMGQ	474
QY	481	QIGSYFSGSELCPDTPDRGTTDVLVAAPMFLGPONKTEKRYVYVVLVQOGLTLTQGLTQ	540
DB	475	QIGSYFSGSEITSVSDIDGGVDVLLVGAAPMFN--EGRERKQYVEL--RQNFVYNGTLK	532
QY	541	PEPP--QDARFPGMALPDLDNODGFADVAVGAFLPDEHGQALYVHGTSQGRPPAPRI	599
DB	533	DSHSYQNNRPFSSISAVSDLNODSINDVYVGAFLPDEHNAGALYIHGGRGSLTKPKQRI	592
QY	600	AAASMPHALSYFGRSVDRDLDDGDDLVAVGAAGAAIILSSRPVHLSLEVTPOAI	659
DB	593	TASELATGLQYFGCSIHGQLDLNEDGLIDLAVGALGNVILMSRVVQIINSLHPEPSKI	652
QY	660	SVYQDCRRRGCEAVCLTALCLFOVTSRTPGRMHQFTMRFTASLDEWTAGARAAFDGSG	719
DB	653	NIFHRDCRRSGRDATCLAAFLCFEPIFLAAPHOTTGIRVATWDEKRYTPRAHLDGCG	712
QY	720	QRLSPRRLISGVNVTCEQLHNVLDTSQDYLRPVALTVPALDNTTKGPVLENGSPSI	779
DB	713	DFETRAVAVLSSGQELCRINHVLDITADYKAPVFSVYSLEDDPH--GPMULDGMPITL	771
QY	780	QKLVFESKDCGPDNECVTDVLQVNMMDI-----RGSRK-----APFVY	817
DB	772	RVSVPFMNGCNEDEHCVDIVLDARSDDLPTAMEYQGRVLRKPAQDCSAVTLSFTVTFII	831
QY	818	RGRKRVLVSTLTENRKNNAVTSISITFSRNHLASLTPORESPIKYCEAPPS--AAR	875
DB	832	ESTORVAVVEATLENRGNNAVSTVINISQSANLQFASIIQKEDSGSIECVNERLQKQ	891
QY	876	LCSVGHAPVQCAKATFLLEPFSSSLLSQVFGKLTASSLSLENGTLQENTAGTASVI	935
DB	892	VCVNSVPPFRAKAKAFRLDFEFSKSIPLHHELEIAGSDSNESTKEDVAALRFL	951
QY	936	QYEPHLLPSSSESTLRYEVHPTLP--VGPGEKTKTLRYQNLGCVYVSGIILSALLPA	993
DB	952	KYEDAVLFTRSSSLSHVEVXKNSLSERDYGCPFSSCFIRIONGLFIHGMKITTPI	1011
QY	994	VAHGNVYLSLSQVITNNASCTIV	1016
DB	1012	ATRSGRNRLKLRDPLTDEGHLV	1034

Wed Jul 16 08:48:31 2003

us-09-647-544-2.open.rag

Page 18

Search completed: July 16, 2003, 07:52:04  
Job time : 135.667 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 07:55:38 ; Search time 84.5295 Seconds

(without alignments)  
1639.577 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106

Sequence: 1 MELPVTHTLFLPLVFLTGLC.....GFAHKKIPEEKREKLEQ 1167

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 segs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.dep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.dep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.dep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.dep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.dep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.dep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.dep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.dep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.dep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.dep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.dep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.dep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.dep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.dep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2429	39.8	1189	9	US-09-984-130-35
2	2429	39.8	1189	9	US-09-836-353A-35
3	2254	36.9	1034	9	US-09-984-130-43
4	2254	36.9	1034	9	US-09-836-353A-43
5	1887	30.9	1151	9	US-09-984-130-103
6	1887	30.9	1151	9	US-09-836-353A-103
7	1756	28.8	1181	9	US-10-160-354-2
8	1727.5	28.3	707	9	US-10-125-540-313
9	1727.5	28.3	707	10	US-09-764-870-313
10	1131.5	18.5	1161	9	US-09-891-943-2
11	1131.5	18.5	1161	10	US-09-350-259-2
12	1126	18.4	1161	9	US-09-891-943-99
13	1126	18.4	1161	10	US-09-350-259-99
14	1124.5	18.4	1161	9	US-09-891-943-55
15	1124.5	18.4	1161	10	US-09-350-259-55
16	1109	18.2	1161	9	US-09-891-943-53
17	1109	18.2	1161	10	US-09-350-259-53
18	1108.5	18.2	1151	9	US-09-891-943-37
19	1108.5	18.2	1151	10	US-09-350-259-37

20	1105	18.1	1170	10	US-09-945-265-2	Sequence 2, Appl
21	1104.5	18.1	688	9	US-09-866-050A-624	Sequence 624, App
22	1103	18.1	1155	9	US-09-891-943-46	Sequence 46, Appl
23	1103	18.1	1155	10	US-09-350-259-46	Sequence 46, Appl
24	1085.5	17.8	1153	9	US-09-902-481A-1	Sequence 1, Appl
25	1085.5	17.8	1153	9	US-09-891-943-3	Sequence 3, Appl
26	1085.5	17.8	1153	9	US-10-144-259-30	Sequence 30, Appl
27	1085.5	17.8	1153	9	US-10-207-655-176	Sequence 176, Appl
28	1085.5	17.8	1153	10	US-09-350-259-3	Sequence 3, Appl
29	1079	17.7	1152	10	US-09-945-265-4	Sequence 4, Appl
30	1065.5	17.5	1137	9	US-09-902-481A-6	Sequence 6, Appl
31	1064.5	17.4	1137	9	US-09-902-481A-5	Sequence 5, Appl
32	1063	17.4	696	9	US-09-866-050A-501	Sequence 501, App
33	1051.5	17.2	1137	9	US-09-902-481A-4	Sequence 4, Appl
34	1042.5	17.1	1137	9	US-09-902-481A-3	Sequence 3, Appl
35	1025	16.8	1163	9	US-09-891-943-4	Sequence 4, Appl
36	1025	16.8	1163	10	US-09-350-259-4	Sequence 4, Appl
37	1006	16.5	1195	9	US-09-805-354-7	Sequence 7, Appl
38	1006	16.5	195	9	US-10-144-259-7	Sequence 7, Appl
39	776	12.7	545	9	US-09-866-050A-500	Sequence 500, App
40	690	11.3	124	9	US-09-978-295A-346	Sequence 346, App
41	690	11.3	124	9	US-09-978-697-346	Sequence 346, App
42	690	11.3	124	9	US-09-978-182A-346	Sequence 346, App
43	690	11.3	124	9	US-09-999-832A-346	Sequence 346, App
44	690	11.3	124	9	US-09-978-189-346	Sequence 346, App
45	690	11.3	124	9	US-10-174-590-154	Sequence 154, App

#### ALIGNMENTS

RESULT 1  
US-09-984-130-35  
Sequence 35, Application US/09984130  
Publication No. US20030055231A1  
GENERAL INFORMATION:  
APPLICANT: NI et al.  
TITLE OF INVENTION: 12 Human Secreted Proteins  
FILE REFERENCE: PFA8992  
CURRENT APPLICATION NUMBER: US/09/984,130  
CURRENT FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: 60/243,792  
PRIOR FILING DATE: 2000-10-30  
PRIOR APPLICATION NUMBER: 09/836,353  
PRIOR FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/198,407  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: PCT/US99/25031  
PRIOR FILING DATE: 1999-10-27  
PRIOR APPLICATION NUMBER: 60/105,971  
PRIOR FILING DATE: 1998-10-28  
NUMBER OF SEQ ID NOS: 149  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 35  
LENGTH: 1189  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-984-130-35  
Query Match  
Best Local Similarity 39.8%; Score 2429; DB 9; Length 1189;  
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;  
1 MELPVTHTLFLPLVFLTGLCSPFNLDENHRLFLFPFAEFGCYSLVGHVGGGQRMLVGA 60  
1 MDLPGLVAVALSIPGFTDTFNMDDTRKPRVIPSRTAFPGYVQOHDSGNKVLVGA 60  
61 PWDGSGDRDQDVYRCFPGAGHNAFCAKGHGADYOLGSSHPAVMHLGMSLLETGCGG 120  
61 PLFTNGYKTDYKCPV---IHGNCTKLNIGRVTLVSERKXNMRLGSLATPKONS 117  
121 FMACAPLWSRAAGSSVFSSGICARVDASFPQGS LAPTAORCPTYMDVIVLDGNSIYP 180

118 FLACSPMLSHBEGSSYYTTGMCGRVNSNFRFSKTVAPALORCQTYMDIVILVDGNSIYIP 177  
181 WSEVOTFLRLVGLKFLIDPEQIOVGLVOGESPVHEMSLQDPFKTEEVVRAAKNLRRREG 240  
178 WVEVGHFLINILKKFPIGRQIOGVVOGEDVHEFLINDYRSVADVVAASHIEQGG 237  
241 RETKTAQAIWVACTEGFSQSHGRPEARLLVVVTGESHGDEELPAALKACEAGRVTRY 300  
238 TETRTAFGIEFARSEAFQK--GGRKAKKVMIVITDGESHSDSDLEKVIQOSERDVTTRY 295  
301 GLAVGHYLRQRDPSSFLREITIASDPERFFNVNTEAALTDIVDALGRIIFGLEGS 360  
296 AAVAILGYNNRGINPFTLNEIYIASDDPKHFNVTDEALKDIDVALGDGIFSGLEGT 355  
361 HAENESSFGLMSQIGFSTHRLKDGLFGVNGAYDMGGSVLMLEGGHRLFPPRMALDEDF 420  
356 N-KNETSFGLEMSQIGFSSHVVEDGVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414  
421 PPALONHAAVLYGYSVSSMLRGRRLFLSGAPRFRRHGRKVIAPOLKKDGAVRVAQSGLOGE 480  
415 PEELKNHGAAYLGTYTVSVSSRQGRVYVAGAPRFNHTGKVIPLTMNNRSLTIHQAMRQ 474  
481 QIGSYFGSELCPIDTRDGTDTVLVAAPMLFGPOKKEGRVYVYLVGOOSLITLOGTQ 540  
475 QIGSYFGSEITSDIDGDTVLVAGAPRYFN--EGREGRKYYVEL--RQNRFFVNGTLK 532  
541 PEPP--QDARFGFAMGALPDLDNODGFADVAVGAPLEDEHOGALYLHGTSQGVAPRPAORI 599  
533 DSHSYQNAFPGSSIASVRDLNODSYNDVVGAPLEDNHAGAIYIFHGRGSLIKTPKORI 592  
600 AAASMPHALSYFGRSVDRDLDDGDDLVDAVGAQGAAILLSRPVHLTPSLEVTPOAI 659  
593 TASELATGLOYFGCSIHQOLDLNEDELIDLAVALGNAAVILMSRPVAVQINASHPEPSKI 652  
660 SVVQDRCRRRQGANCLTAAALCFQVTSKTPGRMWHQFMRTASLDEMTAGARAFDSSG 719  
653 NIFHRDCKRSGRDAVLAFLCTPTPIFLAPHPOTTVGIRYNATMDERRYPRAHLDEGG 712  
720 QRLSPRLRLSVGNVTCQOLHFFHVLDTSDYLRFVALVTTFALDNTKPGPVINEGSPSI 779  
713 DRTNNAVILLSSQGLCERINFPVLDOTADVKKVTPSVESLDEDPH--GPMDDGMPPTLL 771  
780 QKLVPSKDCGPNCECVTDVLVQNMNDI-----RGSRK-----APPV 817  
772 RVSVPMNCGNEDHCVDPDLVDARSDLPJAMEYCORVLRKPAQDCSAYTLSEDTTVFI 831  
818 RGRARRVLYSTLENKENAYNNTSLIISRNHLASLTPORESPIKVCAPAS--AAHR 875  
832 ESTROVAAVEATLENGENAYSTVLNISQSANIQFASLIQKEDSDGSIQVNEERLQKO 891  
876 LCSVGHVFPQTGAQVFLLEFEFSCSLSQVFGKLTASSDSLERNGLQENTAQTSAYI 935  
892 VCNVSVPFRAPRAKAVAFRLDFEFSKSIPLHLLEIELAAGSDSMERSTEDNVAPLRPL 951  
936 QVEPHLLFSSESTLHRYEVHRYGTLP--VGPGPEFTTLRVONLGCYVVSGLIISALLPA 993  
952 KYEADVLFRSSSLSHYEVYLNLSLERVYDGIQPFQICRIQNLGFLPHIGIMMKITIP 1011  
994 VAHAGVYFLSLGQVIN--NASC--IVQNLTEPRGPVHPELQHNRLNGSTTQOQVVR 1049  
1012 ATRSGRRLKLKLDLDEVAANTSQNIWGSTERYPRVE--EDIRRAPOLNHSIDVAVSN 1070  
1050 CHLGOLAKGTEVSVGLRLVHNEFFPRAPKSLTVVSTFELGTEEGSVQLTEASWSS 1109  
1071 CNI--RIVPQOEINFLHGLNMLRSLKALKYKSKIMVNAALQRFHSPIRFEEDSSROI 1129  
QY 1110 LIEVQVTRPILISLMLIGSVLGGILLALLVFCIMKGLFPAHKKIPEEKKEEKLE 1166  
Db 1130 VFEISKQEDMQVPIWIIIVGSTIGLILLALLVLAALKLGFRRSAR----RRREPCHD 1182

; Sequence 35, Application US/09836353A  
; Publication No. US20030129685A1  
; GENERAL INFORMATION:  
; APPLICANT: NI et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: PF489P1  
; CURRENT APPLICATION NUMBER: US/09/836.353A  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,407  
; PRIOR FILING DATE: 2000-04-19  
; PRIOR APPLICATION NUMBER: PCT/US99/25031  
; PRIOR FILING DATE: 1998-10-27  
; PRIOR APPLICATION NUMBER: 60/105,971  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 35  
; LENGTH: 1189  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-836-353A-35

Query Match 39.8%; Score 2429; DB 9; Length 1189;  
Best Local Similarity 43.1%; Pred. No. 1,4e-188;  
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

1 MELPVTNHLFLPLVFLTLGSPFNDEHHPRLFPGPPEAFEGSYLQHVGGGQRMVLVGA 60  
1 MDLPGLVYNAALSLMPGFTDTFNMDTRKPRVIPSRAFCYITQOHDIGNKMLVYGA 60  
61 PMDGGSGRRGDDVYRCVPVGAHNAFCAKGLADYQOLGNSSHPAVNMLGMSLLETDGGC 120  
61 PLETNGYQKTGDGVYKCPV---IHGNCITKLNGRVLTVSVSRKDMRGLSLATNPKNDS 117  
121 EMACAPLMSRACGSSVFSSGICARVDASFOQGSALAPRORCPYMDVIVILVDGNSIYIP 180  
118 FLACSPMLSHBEGSSYYTTGMCGRVNSNFRFSKTVAPALORCQTYMDIVILVDGNSIYIP 177  
181 WSEVOTFLRLVGLKFLIDPEQIOVGLVOGESPVHEMSLQDPFKTEEVVRAAKNLRRREG 240  
178 WVEVGHFLINILKKFPIGRQIOGVVOGEDVHEFLINDYRSVADVVAASHIEQGG 237  
241 RETKTAQAIWVACTEGFSQSHGRPEARLLVVVTGESHGDEELPAALKACEAGRVTRY 300  
238 TETRTAFGIEFARSEAFQK--GGRKAKKVMIVITDGESHSDSDLEKVIQOSERDVTTRY 295  
301 GLAVGHYLRQRDPSSFLREITIASDPERFFNVNTEAALTDIVDALGRIIFGLEGS 360  
296 AAVAILGYNNRGINPFTLNEIYIASDDPKHFNVTDEALKDIDVALGRIIFSLGEGT 355  
361 HAENESSFGLMSQIGFSTHRLKDGLFGVNGAYDMGGSVLMLEGGHRLFPPRMALDEDF 420  
356 N-KNETSFGLEMSQIGFSSHVVEDGVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414  
421 PPALONHAAVLYGYSVSSMLRGRRLFLSGAPRFRRHGRKVIAPOLKKDGAVRVAQSGLOGE 480  
415 PEELKNHGAAYLGTYTVSVSSRQGRVYVAGAPRFNHTGKVIPLTMNNRSLTIHQAMRQ 474  
481 QIGSYFGSELCPIDTRDGTDTVLVAAPMLFGPOKKEGRVYVYLVGOOSLITLOGTQ 540  
475 QIGSYFGSEITSDIDGDTVLVAGAPRYFN--EGREGRKYYVEL--RQNRFFVNGTLK 532  
541 PEPP--QDARFGFAMGALPDLDNODGFADVAVGAPLEDEHOGALYLHGTSQGVAPRPAORI 599  
533 DSHSYQNAFPGSSIASVRDLNODSYNDVVGAPLEDNHAGAIYIFHGRGSLIKTPKORI 592  
600 AAASMPHALSYFGRSVDRDLDDGDDLVDAVGAQGAAILLSRPVHLTPSLEVTPOAI 659  
593 TASELATGLOYFGCSIHQOLDLNEDELIDLAVALGNAAVILMSRPVAVQINASHPEPSKI 652  
660 SVVQDRCRRRQGANCLTAAALCFQVTSKTPGRMWHQFMRTASLDEMTAGARAFDSSG 719  
653 NIFHRDCKRSGRDAVLAFLCTPTPIFLAPHPOTTVGIRYNATMDERRYPRAHLDEGG 712

QY 720 QRLSPRLRLSVGNVTCQQLHFFVLDTSYLRPAVLTAFALDNTTKPGPVINEGSPTSI 779  
Db 723 DFTNRAVALLSSGGELCERINFPVLDLADYVKKVTSVEYSLEDDPH-GPMLDDGWPFTLL 771  
QY 780 QKLVPPSSKCCGPDNECVTLVLYQVNMDI-----RGSRK-----APRVV 817  
Db 772 RVSVPFMNGCNDEHCVPLVLDARSDLPAMEYCORVLRKPAODCSAYTLSFDTTVEFI 831  
QY 818 RGRRRKVLVSTTLLENKENAYNTSLISIFSRNLHLASLTLPQRESPIKVCAPSS-AHAR 875  
Db 832 ESTROKVAVEATLERNGENAYSTVLNISQSANLQFASLTQKEDSDGSIICVNEERLQOK 891  
QY 876 LCSVGHVPTQAKVYTLLEFEFSCSSLSQVFGKLTASSDSLERNGTLQENTAOQTSAYI 935  
Db 892 VCNVSYPPFRKAKAKAFRLDPEFSKSIPLHLEIELAAGSDSNERNSTEDVAPLRPHL 951  
QY 936 QYEPHLLSSSESTLHRYEVHPYGLP--VGPGBEKFTTLRVQNLGCVVSGLIISALLPA 993  
Db 952 KYEADVLFTRSSSLSHYEVKLNSSLERYDGIQPPFCIFRIQNLGIFPIHGIMMKITPI 1011  
QY 994 VAHGNVFLSLSQVITN--NASC-IVQNLTEPPGPVARELOHTNRNGSNTQCOVVR 1049  
Db 1012 ATRSGNRLKLKDFLTDEVAANTSCTIWNSTERYPTVE-EDLRAPQNLHNSDVYSIN 1070  
QY 1050 CHLGQILAKGTEVSGILRLVHNEFFRAKFKSLTVVSTPELGTBEGSVQLTEASRMSHS 1109  
Db 1071 CNL-RLVPGQELNPHLGLMLRLSLKALKYKSMKIMVNALORQFHSPIRFEEDPSROI 1129  
QY 1110 LLEVQVTRPILSLMLIGSVLGLLLALLVFLCMLKGFPAHKKIPEEKEEKEKLE 1166  
Db 1130 VFEISKQEDMQVPIWIVGSTLGLLLALLVIALMLKIGFFRSAR---RRREPGLD 1182

## RESULT 3

US-09-984-130-43  
; Sequence 43, Application US/09984130  
; Publication No. US20030055231A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: P489P2  
; CURRENT APPLICATION NUMBER: US/09/984,130  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/243,792  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 09/836,353  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,407  
; PRIOR FILING DATE: 2000-04-19  
; PRIOR APPLICATION NUMBER: PCT/US99/25031  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR APPLICATION NUMBER: 60/105,971  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 149  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 43  
; LENGTH: 1034  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-130-43

## Query Match

36.9%; Score 2254; DB 9; Length 1034;  
Best Local Similarity 44.5%; Pred No. 2,1e-174;  
Matches 464; Conservative 187; Mismatches 356; Indels 36; Gaps 11;

QY 1 MELPFVTHLFLDVLFLTGLCSPPNLDENHRLFPGPPEAFSGSYVLQHVGGGORMMLVGA 60  
Db 1 MDLPRLGLVAVMALSLMPGFTDTFNMDTKRPRVIPSRTAFPFQYTYQOHDISGNKMLVGA 60  
QY 61 PWDGSGDRGQVYKCPGVGAGNAPCAKHLGDYQLGNSSHPAVNMHLGMSLLETDDGCG 120  
Db 61 PLETNGYOKTGVYKCPV---IHGNCYTKLNLGRVTLISNVSEKDMWRIGLSTATNPKDNS 117

QY 121 FMACAPLMSRACSSVFSFGICARVDASFQDGLAPTAORCPTMYDVIYLDGNSIYIP 180  
Db 118 FLACSPLMHSHEGSSYITTCMSRNSNFRBSKYAPLALOKQOTMIDVIYLDGNSIYIP 177  
QY 181 WSEVQTLRLVYKGLFIDPEQIYQVGYGSGPVHWSLGFRTKEEVRAAKNLSRREG 240  
Db 178 WEVEQHLINILIKRFYIGPQIQVGVQYGEDVHVEFLNDYRSVKDVAAASHIEQGG 237  
QY 241 RETKTAQALMACTGFSQSHGRPEARLNVYVTDGSSHDGEELPAALKACEARVTRY 300  
Db 238 TETRTAFGIEFARSBAFK--GGRGAKKAVIIVTDGSSHPDLKVIQOSEPDNTRY 295  
QY 301 GIATVGHTRRQDDSSFLREIRTAJSPDERFFNVNDEALITDVALGRIFGLRES 360  
Db 296 AVAVLGYNNRKGINETTLNEIKYIASPDDEKHFNVNDEALAKIYDVALGRIFSLBCT 355  
QY 361 HAENESSFGLMSQIGFSTHRLKDGILFGWVGA YDMGGSVYLMBCGHRLPFRMALDEDF 420  
Db 356 N-KNETSFGLMSQIGFSSHVEDGVLLGAVGAYDMNGAVLKETSAKVYIPLRBSYLKEF 414  
QY 421 PPAONHAAYLGYSVSMMLKGRRLFLSAGAPRFRHKKVIAFOLKKGAVRVAOSLOGE 480  
Db 415 PBELNKHAAYLGIVTVSVSSRQGRVYVAGAPRFNHTGKVLFTMHNNRSILTIHQAMRGQ 474  
QY 481 QIGSYFSGSELCPDIDRPGTTDVLVLAAPMFLGPONKTEGRVYVYLVGOQSLTLTQTLQ 540  
Db 475 QIGSYFSGEITISVDIDGIGVTDVLVVGAPMFPN-BGRRGKYYVEL-RQNRFPVNGTLK 532  
QY 541 PEPP-ODARFPFAMGALPDLNQDGFADYVAGAPLEDHOGALYLYHGTOSGVRPPHAPRI 599  
Db 533 DSHSYQNAFSSSIASVADLNQDSYNDVVGAPLEDNAGAIYIFHGRGSLTKPKQRI 592  
QY 600 AAASMPHALSYFGKSVDRDLDDGLVDVAVAGAAIILSSRPYVLTSLSEVTPQAI 659  
Db 593 TASELATGLOQYFGCSIHQDLDLNEDGLIDLAVGALGNVVIIMSREVQVIMNSLHPEPKI 652  
QY 660 SVVORDCRRGOEAVCLTAAALCFQVTSRTPGRMHDQFMRFTASLDEMTAGARAAPDSG 719  
Db 653 NIFHRDCRSGRDATCLAAFLCFPTPIFLAPHPQTTTGVGIRYNATWDEKRYPPRAHLDDCG 712  
QY 720 QRLSPRLRLSVGNVTCQQLHFFVLDTSYLRPAVLTAFALDNTTKPGPVINEGSPTSI 779  
Db 723 DFTNRAVALLSSGGELCERINFPVLDLADYVKKVTSVEYSLEDDPH-GPMLDDGWPFTLL 771  
QY 780 QKLVPPSKCCGPDNECVTLVLYQVNMDI-----RGSRK-----APRVV 817  
Db 772 RVSVPFMNGCNDEHCVPLVLDARSDLPAMEYCORVLRKPAODCSAYTLSFDTTVEFI 831  
QY 818 RGRRRKVLVSTTLLENKENAYNTSLISIFSRNLHLASLTLPQRESPIKVCAPSS-AHAR 875  
Db 832 ESTROKVAVEATLERNGENAYSTVLNISQSANLQFASLTQKEDSDGSIICVNEERLQOK 891  
QY 876 LCSVGHVPTQAKVYTLLEFEFSCSSLSQVFGKLTASSDSLERNGTLQENTAOQTSAYI 935  
Db 892 VCNVSYPPFRKAKAKAFRLDPEFSKSIPLHLEIELAAGSDSNERNSTEDVAPLRPHL 951  
QY 936 QYEPHLLSSSESTLHRYEVHPYGLP--VGPGBEKFTTLRVQNLGCVVSGLIISALLPA 993  
Db 952 KYEADVLFTRSSSLSHYEVKLNSSLERYDGIQPPFCIFRIQNLGIFPIHGIMMKITPI 1011  
QY 994 VAHGNVFLSLSQVITNASCIV 1016  
Db 1012 ATRSGNRLKLKDFLTDEGEHVL 1034

## RESULT 4

US-09-836-353A-43  
; Sequence 43, Application US/09836353A  
; Publication No. US20030129685A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins

```
FILE REFERENCE: PF489P1
CURRENT APPLICATION NUMBER: US/09/836,353A
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 147
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 43
LENGTH: 1034
TYPE: PRT
ORGANISM: Homo sapiens
US-09-836-353A-43
```

```
Query Match      36.9%; Score 2254; DB 9; Length 1034;
Best Local Similarity 44.5%; Pred. No. 2,1e-174;
Matches 464; Conservative 187; Mismatches 356; Indels 36; Gaps 11;
```

```
1 MELPFTLFLPLVFLGLSPFLNDBHRLFPGRPEAEFGYSVLQHVGGGRMLVGA 60
1 MDLPRLVVAWALSLMPGFTDTFNMPTKRPRVIPSRTAFEGYVQOHISGKMLVGA 60
61 PMDPSGDRGDVYRCVPGAHNAPCAKGLGDYQGNSSHPVNMHLSMLETDGDCG 120
61 PLENTGYQKTGDYKCPV---IHGNTKLNLSGRVLTLSNVSERDNRLGLSLATNKDS 117
121 PMAADPLMSRACSSVFPSSGICARVDASFOPQSLAPTORCPTVMDVVIYLDGNSIYP 180
118 FLACSLPMSHECCSSYTTGMCSRANSNFRFSKTVAPALORCQTMIDIVIYLDGNSIYP 177
181 WBSVQFLRLVJLKLIDPEQIOVGIVQGESPVHMSLGDPTKEVYRAAKNLSRRG 240
178 WVEVQFLNLLKRYIIGQOIQVGVQYGEDVHFFHNDKSVKDVVEASHLEQRG 237
241 RETKTQAIMVACTEESQSHGGRPAARLLVVVTGESHDELPALAKACAGRVTV 300
238 TERTIAFGIEFASEAFQK--GGRKAKKMYITTGESHDSLDLEKVIQGESRDVTVY 295
301 GLAVLGHYLRORDPSSFLREIRTIASDPDERFFNVVTDEALTDIVDALGDRIFGLEGS 360
296 AAVAILGYNNRGINPETFLEIKYIASDDPKHFFNVTEBALKDIVDALGRIFGLEGT 355
361 HAENESSFGLMSQIGFSTHRLKDGILFGWVAGYDMGSGVLMLEGGHRLFPRMALEDER 420
356 N-KNETSPGLEMSQIGFSSHVEDGVLLGAVGAYDMNGAVLAKETSAGKVIPLRESTLKEF 414
421 PALQNHAAVLYGSVSMILRGGRLLFLSGAPRRHGRKVIAPOLKKDGAVVAOSLOGE 480
415 PRELKHGAGVLYGTVISVSSRGGRVYVAGAPRFNHTGKVIITFMNNNSLTIHQMRQ 474
481 QIGSYFGSELCPILDRDGTDTVLLVAAPWFLGPNKKEGRVYVYLVGQSLITLQGLTQ 540
475 QIGSYFGSEITSDIDGDVTLVVGAPWYFN--EGREGRKVVYEL--RQNFVYVNGTLK 532
541 PEPP--ODARFPFAMGALLPINDGCFADVAVGAPLEEGHOGALYLYCTGOSVGRPHRQHI 599
533 DSHSYONARFGSSIASVRDINDOSYDVAVVAGAPLENNHAGALYIFHGFSGSLIKTPKQHI 592
600 AAASMPHATSYFGRSVDRGLDLDDGLDVAVAGAOAAIILSSRPVHNTPLSEVTPQHI 659
593 TASELATGQYFGCSIHGQDLNEDGLIDLAVGALGNAYILMSRPVQVQINASHFEPSKI 652
660 SVVQDRCRRGOEAVLTALCFQVTSRTPGRWDHOFYWRFTASLDEWTAGAPAAFGSG 719
653 NIFHRKCRSGSDATLAFLCFTPIFLAPHQTTVIGIRYNAIMDEKRYTPRAHLEDEG 712
720 QLSRPRRLSYNVNVCQDLHFVLDTSQYLARVALYVFPALDNTTKRPVNLDEGPTSI 779
713 DRTNRAVALLSSGQELCERINFVLDTADYVVKPVTFSVEVSLDDPDH--GPMLLDDGWPPTL 771
```

```
QY 780 QKLVFESKDCGPDNECVTDVLQVNMDI-----RGSRK-----APFVY 817
DB 772 RVSVPFMGNCNDHCEVDIVLDARSDLPTAMEYQGRVLKRAQDCSAVTLSPFTVFI 831
QY 818 RGRKRVLVSTLENKKNAYNTSISIFSRNLHSLTPQRESPIKYECAAPS--AHAR 875
DB 832 ESTROVAVEATLENRGENAVSTVINISQANLQFASIIQKEDSDGSIQCVNEERLQK 891
QY 876 LCSVGHVPGQCAKTFPLEPEFSCSLSQVFGKLTASSPSLERNGLTOENTQATSAVI 935
DB 892 VCNVSIYPPFRAKAKAFRLDFEFSKIFLHLIELAAGSDSNERDSTKEDVAPLRPHL 951
QY 936 QYEPHLFSSSESTLHRYEVHPYGLP--VGQPEFKTLRVQNLCCYVSGLIISALLPA 993
DB 952 KYEADVLFTRSSLSLHYEVKLNSSLERYDGLGPPFSCIFRIQNLGLPFIHGMKITPI 1011
QY 994 VAHGNYFLSLSQVITNNASCIV 1016
DB 1012 ATRSGNRLKJLDFLTDGEHVL 1034
```

## RESULT 5

```
US-09-984-130-103
Sequence 103, Application US/09984130
Publication No. US2003005231A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: PF489P2
CURRENT APPLICATION NUMBER: US/09/984,130
CURRENT FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: 60/243,792
PRIOR FILING DATE: 2000-10-30
PRIOR APPLICATION NUMBER: 09/836,353
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
NUMBER OF SEQ ID NOS: 149
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 103
LENGTH: 1151
TYPE: PRT
ORGANISM: Homo sapiens
US-09-984-130-103
```

```
Query Match      30.9%; Score 1887; DB 9; Length 1151;
Best Local Similarity 36.3%; Pred. No. 1.9e-144;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;
```

```
23 ENLDEHRLFPGRPEAEFGYSVLQHVGGGRMLVGAIPMDPSGDRGDVYRCVGAH 82
1 FNVDAKNMTEFSGPVEDFGYTVQYENEBEKVILGSLVLPQPKRAGDVYKCPVGRGE 60
QY 83 NARCAKHLGDYQGNSSHPAV-----NMHLGMSLLETDGSGFMAKPLMSRACSSVF 137
DB 61 SLPCVKLDP---VNTSIPNTEYKEMWTFG--STLVNPNGGFLACPLAVYRGHLHY 115
QY 138 SSGICARVDASFOPQSLAPTORCPTVMDVVIYLDGNSIYPMSEVOTFLRLVGLFI 197
DB 116 TTGICSDVSPTFOVYVNSIAP--VQESCTOLDIVIYLDGNSIYPMDSVTAFLNDLLKMDI 174
QY 198 DPEQIOVLQYQGESPVHMSLGDFTRYEBVRAAKNISRRBGRKTAQAIMVACTGEF 257
DB 175 GKQIOVQIVQYGENVTHEFNLNKYSSTEEVILVAAKKIVQKGGGQTMATLGTARKEAF 234
QY 258 SOSHGREGAARLLVVVTGESHDEELPALKKACEARVYRIGAVLGHYLRORDSS 317
DB 235 TEARGARGVKKVMVITVDGSHDHRLKQVIOQCEDENIQRFSTAILGYNRGNLSTEK 294
```

QY 318 FLEIRTTASDPERFENVNTEBALTDVIALGDRIFGLESGHAEENESSFGLENSQIOF 377  
DB 295 FVEBEIKSIASEPTKEKFFNVSDDELAVITVKTGRIIFALEATADQASAFEMENSQIOF 354  
QY 378 STHRLKDGILFGWVGAYDMGGSVLMLEGGHRLFPRMALEDEFPRLALONHAAYLGYSVS 437  
DB 355 SAHYSQDMWMLGAVGAYDMNGVVMOKASQIIPRNTTNNVSTKKNBEPLASLYGTVNS 414  
QY 438 MLRGGRLFLSGAPFRHGRKVIAFOLKKDGAVRAVQSLQEOIGSYFSELCPLDTR 497  
DB 415 ATASSGDVLYIAGQPRVHTGVYIYRM-EDGNIKILOLSEBOIGSYFGSILTTTIDK 473  
QY 498 DGTTLVLLVAAPMFLGPONKTEGRVYVYLVGOOSLTLTGTOF----- 541  
DB 474 DNTDILLVGA PMVMTKEKEGQKVVYAL-NQTRFQYMSLEPIKOTCCSROHNSCTT 532  
QY 542 ---EPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDHOGALYLYHGTQSVRPHAPOR 598  
DB 533 ENKNEPCGARFGTALAAVADLMDGFNDIVIGAPLEDHOGAVYIYHSGKTIKREYAR 592  
QY 599 IAAAMPALSYFGSVGRDLDDGDDLVDAVAGAGAILLSRPYHILTPSLEVTPOA 658  
DB 593 IPSGDDGKTLKFGGSIHGBMDLNGDLTVITIGLGAALFWSRDAVAVKXTMNFEPK 652  
QY 659 ISVORDCRRGOEAVCLTAALCFQVTSRTGPMWHDQFYMRFTASLDEWTAGARAAPDGS 718  
DB 653 VNIQKNCHMEGKETVCINATVCFEYKLSKEDTIEADLOYRVLDSLRQISRFSGT 712  
QY 719 GORLSPRLRLSVGNVTEQOLHFNVLDTSDYLRPALVYTFALDNTTK--GPVINEGSP 776  
DB 713 QORRVOR--NITVRKSECTKHSFYMLDKGDFODSVRIITLDF--NLTPENGSPVLDLSP 767  
QY 777 TSIOKLVPFSKDCGPNDECVTDLVLOVMDIRGSRKAPFVNGGRKVLVSTLENRKEN 836  
DB 768 NSVHEIYIPAKDCGKKEKICISDLSHV---ATTEKDLIYRSQNDKRVSLIVKNTDS 823  
QY 837 AYNTSLSTIFSRNLHLASLTPORESPIKVECAPSAHARLCSVGHVPFOTGAKVTFLEF 896  
DB 824 AYNTRTIYHSPNLVFSGI---EALQKOSC--ESNHNTCKVGPFLRGEWTFKILF 877  
QY 897 ESSCSLSQVFGKLTASDSLEENGTLOENATQTSAYIQYEPHLLFSESTLHRYENP 956  
DB 878 QNTYVLMENVTIYSAISDSSEPPETLSDNVNISIPIKVEGLOFYSASAEYHISIA 937  
QY 957 YGTLF-----VGPGEFKTLRVONLGCYVSGILISALLPAVAHGN--YFLSLSOV 1007  
DB 938 NETVEVINSTEDIGNEINIFYLIRKSGSFMPBELKLSIFPMNMSGPVLYPTGLSS- 996  
QY 1008 ITNNAACIVONLTP-----DGPVHPELOHTNRNLNGSNTQOQVVRCHLOLAKTEV 1061  
DB 997 -SENANCRPHIFEDPFSINGKMTSTDLKRGITLDCNTCKPATITCNLTS-SDISGV 1054  
QY 1062 SVGLRLVHNEFFRAKFKSLVSVTFELGTBEGSLDOTEASRMSSESLLEVOT-RPL 1120  
DB 1055 NVSL--ILMKPTFIKSYFSLMLTIRGELRSENAS-LVSSSNQKELALQISKGLPGR 1111  
QY 1121 ISLMLTIGSVLGLLLALLLVFCMLWLGFEFAHKKIPEEKREE 1163  
DB 1112 VFLWVILLSAFAGLLMLLMLLILALMKIGF---KRLPKKMEK 1151

RESULT 6  
US-09-836-353A-103  
; Sequence 103, Application US/09836353A  
; Publication No. US20030129685A1  
; GENERAL INFORMATION:  
; APPLICANT: NI et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: PF489P1  
; CURRENT APPLICATION NUMBER: US/09/836,353A  
; CURRENT FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,407  
; PRIOR FILING DATE: 2000-04-19

QY 318 FLEIRTTASDPERFENVNTEBALTDVIALGDRIFGLESGHAEENESSFGLENSQIOF 377  
DB 295 FVEBEIKSIASEPTKEKFFNVSDDELAVITVKTGRIIFALEATADQASAFEMENSQIOF 354  
QY 378 STHRLKDGILFGWVGAYDMGGSVLMLEGGHRLFPRMALEDEFPRLALONHAAYLGYSVS 437  
DB 355 SAHYSQDMWMLGAVGAYDMNGVVMOKASQIIPRNTTNNVSTKKNBEPLASLYGTVNS 414  
QY 438 MLRGGRLFLSGAPFRHGRKVIAFOLKKDGAVRAVQSLQEOIGSYFSELCPLDTR 497  
DB 415 ATASSGDVLYIAGQPRVHTGVYIYRM-EDGNIKILOLSEBOIGSYFGSILTTTIDK 473  
QY 498 DGTTLVLLVAAPMFLGPONKTEGRVYVYLVGOOSLTLTGTOF----- 541  
DB 474 DNTDILLVGA PMVMTKEKEGQKVVYAL-NQTRFQYMSLEPIKOTCCSROHNSCTT 532  
QY 542 ---EPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDHOGALYLYHGTQSVRPHAPOR 598  
DB 533 ENKNEPCGARFGTALAAVADLMDGFNDIVIGAPLEDHOGAVYIYHSGKTIKREYAR 592  
QY 599 IAAAMPALSYFGSVGRDLDDGDDLVDAVAGAGAILLSRPYHILTPSLEVTPOA 658  
DB 593 IPSGDDGKTLKFGGSIHGBMDLNGDLTVITIGLGAALFWSRDAVAVKXTMNFEPK 652  
QY 659 ISVORDCRRGOEAVCLTAALCFQVTSRTGPMWHDQFYMRFTASLDEWTAGARAAPDGS 718  
DB 653 VNIQKNCHMEGKETVCINATVCFEYKLSKEDTIEADLOYRVLDSLRQISRFSGT 712  
QY 719 GORLSPRLRLSVGNVTEQOLHFNVLDTSDYLRPALVYTFALDNTTK--GPVINEGSP 776  
DB 713 QORRVOR--NITVRKSECTKHSFYMLDKGDFODSVRIITLDF--NLTPENGSPVLDLSP 767  
QY 777 TSIOKLVPFSKDCGPNDECVTDLVLOVMDIRGSRKAPFVNGGRKVLVSTLENRKEN 836  
DB 768 NSVHEIYIPAKDCGKKEKICISDLSHV---ATTEKDLIYRSQNDKRVSLIVKNTDS 823  
QY 837 AYNTSLSTIFSRNLHLASLTPORESPIKVECAPSAHARLCSVGHVPFOTGAKVTFLEF 896

PRIOR APPLICATION NUMBER: PCT/US99/25031  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR APPLICATION NUMBER: 60/105,971  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 103  
; LENGTH: 1151  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-836-353A-103

Query Match 30.9%; Score 1887; DB 9; Length 1151;  
Best Local Similarity 36.3%; Pred. No. 1.9e-144;  
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;

QY 23 FVLDHHRRLFPPEARFEGSYVLOHVGCGRMLVGA PMWGPSDRGRGDVYRCVGGAN 82  
DB 1 FNVDRKNSMTSPSGPVEDMFGYTVQOYENEGKMWLIGSPLVGQPKRTGDDVYKCPVGRGE 60  
QY 83 NAPCAKGLGIDYOLGNSHPAV-----NMHLGMSLETDGDDGFMACAPLMSRACSSSVF 137  
DB 61 SLPCVKLDP-----VNTSIPNTEVKEKMTFG-STLVNPNNGFLACPLIAYRCGLHY 115  
QY 138 SSGICARVDASFQOGLAPTAORCPYMDVIVYLDGNSIYPMSEVOTFLRLVYKLEI 197  
DB 116 TTGICSDVSPFQOVNSIAP--VQECSTQDLIYIVLDGNSIYPMDSVTAFLNDLKRMDI 174  
QY 198 DPEQOVLVOYGESEPVHMSLGDRTEEVYRAKANSRREGRETAKPAQAIMACTGCF 257  
DB 175 GPKQOVIYVOGEVNTHEFNLNKYSSTEEVALVAAKKIVORGRQTMALGTDTRKKA 234  
QY 258 SOSHGREGAARLVLVVDGESHDEELPALKACAGVTRVYGAIVGHYLRQRDPSS 317  
DB 235 TEARARAGVKKWIVYVDGESHDRKAKYIQDCEDENIQFSAIIGSYNRGULSTEB 294  
QY 318 FLEIRTTASDPERFENVNTEBALTDVIALGDRIFGLESGHAEENESSFGLENSQIOF 377  
DB 295 FVEBEIKSIASEPTKEKFFNVSDDELAVITVKTGRIIFALEATADQASAFEMENSQIOF 354  
QY 378 STHRLKDGILFGWVGAYDMGGSVLMLEGGHRLFPRMALEDEFPRLALONHAAYLGYSVS 437  
DB 355 SAHYSQDMWMLGAVGAYDMNGVVMOKASQIIPRNTTNNVSTKKNBEPLASLYGTVNS 414  
QY 438 MLRGGRLFLSGAPFRHGRKVIAFOLKKDGAVRAVQSLQEOIGSYFSELCPLDTR 497  
DB 415 ATASSGDVLYIAGQPRVHTGVYIYRM-EDGNIKILOLSEBOIGSYFGSILTTTIDK 473  
QY 498 DGTTLVLLVAAPMFLGPONKTEGRVYVYLVGOOSLTLTGTOF----- 541  
DB 474 DNTDILLVGA PMVMTKEKEGQKVVYAL-NQTRFQYMSLEPIKOTCCSROHNSCTT 532  
QY 542 ---EPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDHOGALYLYHGTQSVRPHAPOR 598  
DB 533 ENKNEPCGARFGTALAAVADLMDGFNDIVIGAPLEDHOGAVYIYHSGKTIKREYAR 592  
QY 599 IAAAMPALSYFGSVGRDLDDGDDLVDAVAGAGAILLSRPYHILTPSLEVTPOA 658  
DB 593 IPSGDDGKTLKFGGSIHGBMDLNGDLTVITIGLGAALFWSRDAVAVKXTMNFEPK 652  
QY 659 ISVORDCRRGOEAVCLTAALCFQVTSRTGPMWHDQFYMRFTASLDEWTAGARAAPDGS 718  
DB 653 VNIQKNCHMEGKETVCINATVCFEYKLSKEDTIEADLOYRVLDSLRQISRFSGT 712  
QY 719 GORLSPRLRLSVGNVTEQOLHFNVLDTSDYLRPALVYTFALDNTTK--GPVINEGSP 776  
DB 713 QORRVOR--NITVRKSECTKHSFYMLDKGDFODSVRIITLDF--NLTPENGSPVLDLSP 767  
QY 777 TSIOKLVPFSKDCGPNDECVTDLVLOVMDIRGSRKAPFVNGGRKVLVSTLENRKEN 836  
DB 768 NSVHEIYIPAKDCGKKEKICISDLSHV---ATTEKDLIYRSQNDKRVSLIVKNTDS 823  
QY 837 AYNTSLSTIFSRNLHLASLTPORESPIKVECAPSAHARLCSVGHVPFOTGAKVTFLEF 896

```

Db      824  AYNRTIYHISNVLVPSGI-----EAIQDSC--ESNHNITCKVGFPLIRGEMVTFKLLF  877
OY      897  EFSGSSLSQVFGKLTASDSIERNGTLOENTAOQTSAYIQYEPHLLFSSESTLAREVHP  956
Db      878  QFNSTYLMENVNVIYLSATSDSEEPETLSDNVNINISVPKVEVGLQFSSASEYHISIAA  937
OY      957  YGTLR-----VGPPEFKTTLRVQNLQYVVSGLISALLPAVHAGN---YFLSTSQV  1007
Db      938  NETPEVINSTEDIDONEINIFYLLIKSSFPPEPEKLISIPFNMTSNGPVLYPTGLSS  996
OY      1008  ITNNASCIYONLTPE-----PGPVEHBELOHTRLNGSNTOCOVVRCHLQOLAKTEV  1061
Db      997  -SENANCPHIFEDFPSSINGCKMSTSDHLKRGITLDONCTCKFATITCNULTS-SDISQV  1054
OY      1062  SVGLIRLVHNEFFRRPAKFSKLTVTSFELGTEBGSVLOLTEARSBSLLEVYQT-RPIL  1120
Db      1055  NVSLT--ILMKPTFIKSYSSSLNLTTRGELRSENAS-LVLSASNQRELAIOISKDGLPGR  1111
OY      1121  ISIMLLIGSVLQGLLLALLVPCMLKLPFAHKRIPEEKREE  1163
Db      1112  VPLWVITLSAPAGLLMLMLITLAIKMGF--KRLPKKKEK  1151

```

```

RESULT 7
US-10-160-354-2
; Sequence 2, Application US/10160354
; Publication No. US20030013107A1
; GENERAL INFORMATION:
; APPLICANT: Chu, Peter
; APPLICANT: Li, Congfen
; APPLICANT: Liao, X1 Charlene
; APPLICANT: Pardo, Jorge
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Alpha 2 Integrin: Modulators of Lymphocyte Activation
; FILE REFERENCE: 021044-001110US
; CURRENT APPLICATION NUMBER: US/10/160,354
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/296,819
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1181
; TYPE: PRN
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: wild type alpha 2 integrin, CD49b
US-10-160-354-2

```

Query Match	28.8%	Score 1756;	DB 91	Length 1181;
Best Local Similarity	34.3%	Pred. No. 9,2e-134;		
Matches	418;	Conservative 219;	Mismatches 455;	Indels 126;
			Gaps	31;
QY	11	LPVLVFTGL-----GSPFNLDENHPPLPGPPPAEFGSYVLQHVGGGGRWMLVGABWD	63	
Db	11	LPRLVLVALSGQILNCCLAAYVNGLPBEAKIFSGPSSGQGYAVQGFINKFGMWLLVGGSPWS	70	
QY	64	GPSGGRGDVYRCPCVPGAGANP.PCAKGHL-GDYOLANSHPVNMHLGSLLETDDGGFM	122	
Db	71	GFPENRMDDYKCPV-DLSTATCELTNIQTSTISINVEEMKTNMNSGLILRMNCTGGFL	129	
QY	123	ACAPLMSFACGSSVFFSSGICARVDS.FDPQGS.LAPTAORCPTYMDVYIVLDGNSIYWS	182	
Db	130	TCGFLMAQCCNQYTTTVCSDISPFDLSASFSPATPCSLLDVVVVCCDSNSIYWD	189	
QY	183	EVQTFELRLVYKFLFDPEQIQVGLVOYQES.PHEHSLD.FPTKEEVVAAKULSRERE	242	
Db	190	AVKNPFLKFEVGGDLDGPTKTQVGLIQYANNRRVYVNNLTYYTKKEMLIATSOTQOYGDLD	249	
QY	243	TKTQALINVACTEGFSQSHGGRPEAKRLVVYTTDSESHDGEELPAALKAACEAGVTRYGI	302	
Db	250	TNTGTCALQIYARKYVAASAAGRRSRATKYMMVVVTTDGEESHGMLKAAVDDCHNDHILRGI	309	

```

QY 303 VILGHYLRQRPSSFLREFTTISDPREFEYVNTBEALTDVLADGBIFGLGSHA 362
Db 310 AVLGUNLNNAIDTKNLKELIKALISIPERYFVNSDAALEAKGTLGBOIFISEG-V 368
QY 363 ENESSFGLNMQISFST--HRLKGLIFGWAQVADMGSVLM--DEGHRLEPPMALEDE 419
Db 369 QCGNPNQWMSQVGFSAQYSSQNDILMLGAVGAGMSGTIYQKXSHGLIFP-----KQA 423
QY 420 FPPALQ--NHAAYIGYSSVSMRLNGRRLFISGAPREFHRKGVTAFOLKQDAVVAOSL 477
Db 424 FDOILJDRNHSYSGYVA--ISTGESTHFVAGAPRANYTQIYLVSNENGINITVIOAH 482
QY 478 QGEOIGSFGEICLPJDRDGTDDVLVYAPMLGQNKETGVVVY-----LVGOOSL 532
Db 483 RGGDIGSFSGVLCSDVDKDTIDVLVYAPMWSMDKKEGVVYLTIKKGLIGQOF 542
QY 533 LTLQGLQPEPPODAREFGAMGALPDLINODAFADAVAGAPLEDBHOGALYLHYGTOSGVR 592
Db 543 --LEG---PEBENTRGSAIALSDINDMGDFNDVYGSPLFNONSAGAVITYNHQIGIR 597
QY 593 PHAPQIAAA--SMPHALSYFGSVDGRLLDGDLDVAVAGQAAIILSSRPVYHLP 650
Db 598 TKYSQKILGSGARSHLOQYGRFLSDGDLNGDSITDVSIGAGQVQLWQSIAVAI 657
QY 651 SLEVTPOALISVYODCCRGRGEAVCLTALCFQYTSRTPGWHQOFYMRFTASIDEMWAG 710
Db 658 EASFTEPKITLVNNAQ-----ILKCFSAKFR--PKQNNQVAIVNYITLDA--- 704
QY 711 ARAAFDGGGQRLSPRL-----RLSYGVN-----TCEOLHFVPLDTSDLREVALVT 758
Db 705 ----DEGSSVTSRGLFKENNECLOKNNVYVNOAQSCPEHIIYIOESPDVVNSLDLRVD 759
QY 759 FALDNTTKPG--PVLNBSPTSIQKLVFPSKDCGPNECVTDVLVQYNNMIDRGRKAPV 816
Db 760 ISLEN---PGTSPALEAVSEYAKVFSIPFHDKCGEDGICISDLVDVLR-QIPAAQEOFPFI 815
QY 817 VRGRRKVYSTLENNKENNAYNYSLSIIFSRNLHLSLTPQRSPI---KVEC-APSA 872
Db 816 VSNONKRTFESVTLKNNRESAYNVGIADVEENLFPASF---LPADGTEVTCQVAAO 871
QY 873 HARLCSVGHAPFOTGAKVTLFLEEFSCSILSQVFCKLTASSDSLEBNGTLQENTACTS 932
Db 872 KSVACDVGYALRKQOVTFTINDPULQNLQNASISFQALSSQENKA--DNLVNLK 929
QY 933 AYIOYEPHLLFSSESTLHRYEYHDPYLP-----VGPBEFKTLRAYONIGCVVSGLI 986
Db 930 IPLLTYDAEHLTRSTINIFEISDQNVPSIVHSFEDVGPXFISLKV--TTGSVPVSMAT 988
QY 987 ISALLPVAHNGNFVLSQVITNNACIQONLTP-----GPPVHEELOHNRLL 1038
Db 989 VIIHPIPYTERKENLWYLTGVQTDKADISCNADINPLKIGOTSSSVSFKESEFRHTEL 1048
QY 1039 NGSNTQCOVNRCHGOLAKTEVSVGLRLVTHNEFFPRAKFSLTIVSTFELGEEGCVL 1098
Db 1049 NCRTRASCSTNYTCLKDVHMKGEYVNVYTRIMNGTFASSITQYVQLTAAAEINYNBEIY 1108
QY 1099 QLTASRWSBSLLEVQOTRPLISLMT-----LIGSVLGGLLILALVFCW 1145
Db 1109 -----VIDNVTYIPLIMTKPDEKAEVFTGVIIGIILGILLILALVALW 1154
QY 1146 KLGFPAK-----KIPBE 1158
Db 1155 KLGFKKRYEKMTNPDE 1172

```

RESULT 8  
 US-10-125-540-313  
 ; Sequence 313, Application US/10125540  
 ; Publication No. US2003005875A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

? FILE REFERENCE: PT214C1
? CURRENT APPLICATION NUMBER: US/10/125,540
? CURRENT FILING DATE: 2002-04-19
? Priority Application removed - See File Wrapper or Palm
? NUMBER OF SEQ ID NOS: 646
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 313
?
? LENGTH: 707
?
? TYPE: PRT
?
? ORGANISM: Homo sapiens
?
US-10-125-540-313

```

Query Match	28.3%	Score 1727.5	DB 9	Length 707
Best Local Similarity	49.8%	Pred. No. 8.6e-133		
Matches 340, Conservative	125, Mismatches 209		Indels 9	Gaps 6

Qy	I	MELPVTNHLFLPLVFLTGICSPFNLDENHRLFCGPBEAFGYSVLQHVGGQRMVLVGA	60
Db	17	MDLPFGLVAVAMALSLMPGFTDTFNMNDRKPRVYFGSRTAFGTYVQOHDISGNKMLVGA	76
Qy	61	PWDGSGRRGRGVNCPVCGAMNACAKHLDYOLGNSHPAVNMHLSMLLETDGDSG	120
Db	77	PLETNGYOKTGVNVCPU--IHNGCTKNILNRVTLNSVSEKDNMRJGLSLATNPKDMS	133
Qy	121	FMACAPLMSRACGSSVSFSSGICARVDASFOPOGSLAPYQRCPTVMYDVIYLDGNSIYP	180
Db	134	FLACGPLMSHEGSSYTTGCMCSRNSNRPSTKATYAPALQRCQTYMDIYIVLDGNSIYP	193
Qy	181	WSEVQFLRLVYKGLFIDPEQIOVGLVOYGESPVHWSLGDPTKEEVVRAAKNLSRREG	240
Db	194	WVEVQHFLINLIKKEYIYGQIOVGVOYGEIVHFEHLINDYRSVADVEAASHIREGG	253
Qy	241	RETKRQALIMVACTGFSQSHGCRPREAALLVVTVDGESHDEELPALKACEAGRVTRY	300
Db	254	TETRTAFGIEFRASFAFOK--GGRGAKKVMVITDGESHDPDLEKVIQOSERDVTRY	311
Qy	301	GIAYGHYLRORDESSFLREIRTYLSDPDERFFVNTDEALTPIDVALGDIRFGLGSS	360
Db	312	AVAVLYGNRRKQINETFLNELIKYIASDDDHGFVNTDEALKDIDVALGDIRFSLEGT	371
Qy	361	HAENESSFGLEMSQIGFSTRHLKDGILFCGVAGADWGSVLMLEGGHRLFPRMALDEEF	420
Db	372	N-KNETSPGLEMSQIGFSSHVEDGVTLGAVAGADWNGAVLKEYSAGKVIPLRESYLKEF	430
Qy	421	PPALONHAAYLYGYSVSMILRGGRRLFLSGARFRHKGVIYAFOLKGDGAVRVAOSLOGE	480
Db	431	PEELKNHAYLYGITYTISWSSRQGRVYVAGARFNHTGKVIILFTMHNNRSLTIHQMRGO	490
Qy	481	QIGSYFSGELCPLDTRDGGTTDVLLVAAFMFLGPOUKERGVYVYLVOGQSLLTLOGTQ	540
Db	491	QIGSYFSGELITSDIDGDDVTDVLLVAGAMYRN--EGREKGKYYVEL--KQNLFYVNGTLK	548
Qy	541	PEPP--QDARFGFAMALPDINODGFADVAVGARPLEBDGHGALUYLHNGTOSGVRPHPAQR	599
Db	549	DSHSYQNNARFGSSIASVRDINDSDSYNDVVVGARPLEBNHAGALYIFHGFPGSLTKTPKQXI	608
Qy	600	AAASHPHALSTFGRSVDGRLDGDGLDVAVAGACALITLSSRPVHLTPSLLEVTPQAI	659
Db	609	TASELATQLOYFGCSIHGQLDJNEBGLIDLAVAGALNAVILMSRPVQINASLHFEPSKI	668
Qy	660	SVVQDRGCRRGGEAVCLTALCF	682
Db	669	NIFHRDCKRSGDATCLAAFLCF	691

```

RESULT 9
US-09-764-870-313
; Sequence 313, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.,
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT14

```

```

; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-764-870-313

```

Query Match	28.3%	Score 1727.5	DB 10	Length 707
Best Local Similarity	49.8%	Pred. No. 8.6e-132		
Matches 340	Conservative 125	Mismatches 209	Indels 9	Gaps 6

[illegible]

RESULT 10  
US-09-691-943--2  
Sequence 2, Application US/09691943  
Publication No. US2003007278A1  
GENERAL INFORMATION  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vliet, Monica  
TITLE OF INVENTION: NO. US2003007278A1 Human 2  
FILE REFERENCE: 27866/35004



```

CURRENT APPLICATION NUMBER: US/09/891,943
CURRENT FILING DATE: 2001-06-26
PRIOR APPLICATION NUMBER: 09/193,043
PRIOR FILING DATE: 1998-11-16
PRIOR FILING DATE: 1994-08-05
PRIOR FILING DATE: 1994-08-05
PRIOR FILING DATE: 1994-08-05
PRIOR FILING DATE: 1994-12-21
PRIOR FILING DATE: 1997-10-03
PRIOR FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 1161
TYPE: PRT
ORGANISM: Homo sapiens
US-09-891-943-2

Query Match      18.5%; Score 1131.5; DB 9; Length 1161;
Best Local Similarity 29.4%; Pred. No. 6,6e-83;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDHHRFLRPGPEAFEGSYVLQVGGGQRMMLVGAIPWDGSGDRGRDVRCPVGAH 82
DB 17 FNLDVEEPTIFQ-EDAGFGGOSVVOF--GGSF-LVVGAPLEVVAAVQGRLYDC---AAA 69
QY 83 NAPCAKGIHGDYQJGNSHPAVNMHIGMSLLETGDDGGFMAACAFLMSRACGSSVFSGIC 142
DB 70 TGMCOPIPL-----HIREPAVNMISGLTLASTNGSRLLACGPTLHRVCGENSYSKSGC 123
QY 143 ARVDASFQPGSLAFTAQRCP-TYMDVVIYLDGNSI--YPMSEVQTLRLRVGKLFIDP 199
DB 124 LLLGSRWEIITQVDPATPECPHOEMDIVFLIDSGSIDQNDFQMGFVOAVWGQ--FEG 181
QY 200 EQIQVULVQYGESEPHNEMSGDRRTKEEVYRAKULSREGRRTKQAQIMACTGCFQ 259
DB 182 TDTLFLMCOYSMLKIHFTFTQSPSQSLVDPIVLQKGL-TFATGILTVVTLFHH 240
QY 260 SHGGRPEARLAVVVTGESH--DGEELPALKACBAARVTRGIVAVLHVRQDPSPSF 318
DB 241 KNGARSAKKILVITDCKYKPLEYSDVIPAQAKGIIIRVAGV--GHAF---QGPFA- 295
QY 319 LREIRTIASDPDERFFFNVTDEALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGFS 378
DB 296 ROELNTISSAPPODHFVKYDNFALGSIQKLOEKIYAVEGTQSRASSFOHEMSQEGFS 355
QY 379 TIRLKGIILFGWGAVDWGSVLMLEGRHLRPPRMALDEFFPALQNH-----AY 430
DB 356 TALTMGDLFLGAVGFSW-----SGCAFLYPPNMS-----PTFINMSQENVMDRSDY 402
QY 431 LGYSVSMLLRGGRRLFLSGAPRRHARGKVIAF-QLKKGDAVVAOSLQEOIGSYFGSE 489
DB 403 LGYSTLALMKQONVL--GAPRYCHTKAVITQVSRQW--RKKAEVITGTOGIFGAS 459
QY 490 LCPDLTDRTDVTDLVLAAPMLGPNKETGRVYVVL--GQGSLLTLQGLQPEPPD- 546
DB 460 LCSVDVDSGSTDLILIGAPHY--EQRGGQVSVCLPRGQGVQOCDAVLRGEGHWP 517
QY 547 ARFGFAMGLPDLNOCGFADVANGARLEBDHOGALTYHC--TOSGVRHPRAPORIAASMP 605
DB 518 GRRGALTYLGVDNNEKLIDVALGARGEBENRGAVVLLFGASGSGISPSHOSIASQSLS 577
QY 606 HALSYGRSVGDGLDDGLVDVAVAGAGAAALLSSRPVHLTPLEVTTPQALISVQGD 665
DB 578 PRLOYTGQALSGGODITOGDLMDLAVGARGQVILLSLSLVLYKKVGMAMRSPEVEAAVVR 637
QY 666 C-----RRRGEAVCLTAALCFQYTSRTPGRMHQFYWRFTASLDENTAGARAAFDGS 718
DB 638 CMEKPSALEAGATVCLTIQ-----KSSLDQGLDQSSVRFDLALDPERLSRAIFNET 692
QY 719 GQRLSRRLRLSLGANTVCEQLHNVLD--TSDYLRPALVTVTALDNTTTPG-----PVN 772
DB 693 KQPTLTRRRTLGIG-THCETLKLPLDCEVEDVVSPIILHNSFLVREPIPSQNLRPVLA 751
```

```

QY 773 EGSPTSIQKLVFPSKDCGPDNECVTDLVLOVMMDIRGSRKAPFVVGRRKVLVSTLTLEN 832
DB 752 VSSQDLFTASLPFEKNCODGCEBD--LGVTLSGSIQT---LVGSSLELNAVITVWN 806
QY 833 RRENAVNTSLSTIFSRNHLASLPDRESP---IKVEC-AAPSAHARL---CSVGHV 883
DB 807 AEDSYGVAVSLYYPAGLSHRSGAQKQPHOSALRLCETVPTDEGRSSRCVSNPI 866
QY 884 FOTGAKVTFLEEFEESSCSLLSQVFGKLTASSDSLERNGTLOENATQTSAYIQYEPHLF 943
DB 867 FHESNGFTVITVDFVSYPATLADRM-LMPASSSNNKASSKATFQLELVKIVVTVI 925
QY 944 SSESTLHREYVAPYCTLPVGPPEFRTTLRYQNLCCYVVSGLIISALLPVAHGNFPLS 1003
DB 926 SNGEESTKY--FNFAITSEKQKWEHRYRVNVLISQRLA-LSINFWPVLLNGVAVMDV 982
QY 1004 LSOVITNNASCTVQNLTEPPGPVPAPELQHTNR--LNGSNTQCVVVRCHGLQAKTE 1060
DB 983 VNEASQSLPCVSEF-----KPPCHSDPLTQISRSPMLDCSIADCLQRCDCVPSPVOE 1037
QY 1061 VSVGLRLVHNEFFRAKFKSLTVSTFELGTEBESVLQLTASRWSSLE-VVQTRPI 1119
DB 1038 LDFTLKGNLSFGWVETLQKVLVSVVAEITFDTSVVSQLEPQEAFMRAQEMVLEDBV 1097
QY 1120 LISLWILIGVGLLTLALVFLCMKLGFF-AHKKIPEEKREE 1163
DB 1098 YNAIPIIMSSVGLALLLALITATLYKLGFFRHHKEMLEDPED 1142

RESULT 11
US-09-350-259-2
Sequence 2, Application US/09350259
Patent No. US20020062008A1
GENERAL INFORMATION:
APPLICANT: Galatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. US20020062008A1 Human 2
FILE REFERENCE: 27866/35004
CURRENT APPLICATION NUMBER: US/09/350,259
EARLIER FILING DATE: 1999-07-08
EARLIER APPLICATION NUMBER: 09/193,043
EARLIER FILING DATE: 1998-11-16
EARLIER APPLICATION NUMBER: 08/173,497
EARLIER FILING DATE: 1993-12-23
EARLIER APPLICATION NUMBER: 08/286,889
EARLIER FILING DATE: 1994-08-05
EARLIER APPLICATION NUMBER: 08/362,652
EARLIER FILING DATE: 1994-12-21
EARLIER APPLICATION NUMBER: 08/943,363
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 1161
TYPE: PRT
ORGANISM: Homo sapiens
US-09-350-259-2

Query Match      18.5%; Score 1131.5; DB 10; Length 1161;
Best Local Similarity 29.4%; Pred. No. 6,6e-83;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDHHRFLRPGPEAFEGSYVLQVGGGQRMMLVGAIPWDGSGDRGRDVRCPVGAH 82
DB 17 FNLDVEEPTIFQ-EDAGFGGOSVVOF--GGSF-LVVGAPLEVVAAVQGRLYDC---AAA 69
QY 83 NAPCAKGIHGDYQJGNSHPAVNMHIGMSLLETGDDGGFMAACAFLMSRACGSSVFSGIC 142
DB 70 TGMCOPIPL-----HIREPAVNMISGLTLASTNGSRLLACGPTLHRVCGENSYSKSGC 123
QY 143 ARVDASFQPGSLAFTAQRCP-TYMDVVIYLDGNSI--YPMSEVQTLRLRVGKLFIDP 199
```



```

Db      124 LLLGSRWEIIQTVPDPATPECPHOEMDIVLIDGSSGSDIONDNFNOMKGFVQAVMGO--FEQ 181
Qy      200 EOIQVGLVOYGESPVHEMSLDGPFRTKEEVBAKKNLSREGETKTAQAINVACTEGESQ 259
Db      182 TDTLFLMQLNSLKHFTFOFRITSPSOOSLVDPIVQJLGL-TFTATGILLVVQTLFPH 240
Qy      260 SHGGRPEARLLVVTGDESH--DGEELPALKACAGATRYGIAVLGHYLRQRDPSSF 318
Db      241 KNGARSAKKILIVITDQKYPLEYSVDIIPAERKAGIIRYALGV-GHAF--QGPTA- 295
Qy      319 LREIRTIASDPDERFFENVTDAAALTDIDALGDRIFGLESHAENESSFGLMSQIGFS 378
Db      296 ROELNTISSAPPODHVFVDNFPALGSIQKQLOEKIYAVEGTQSRASSSFQHEMSQEGFS 355
Qy      379 THRLLKGLIFGNVAGAYDMGGSVLMLEGGHRLPPPRMALEDEFPALONHA-----AY 430
Db      356 TALTMGDLFLGAVGSFSW-----SGAFILYPRNMS-----PFIINMSQENVMDRSDY 402
Qy      431 LGYSVSMULRGGRRLFLSGAPPRHRKGVIAF-QLKQDGAVRVAOSLOGEIGSYFSGE 489
Db      403 LGYSTELALMKGVQNLVL-GAPRYOHTGKAVIFTQVSQW--RKAEVLTGTQISYFGAS 459
Qy      490 LCPIDTRDGTDTVLVLAAPMFLGPONKETGRVYVVLV--GQOSLLTLOGTQOPPEPD- 546
Db      460 LGSVDVSDGSTDLLILIGAPHYY--EOTRGQGVSCPLPRGRQVOMQCDAVLRGEQHPW 517
Qy      547 ARFGFAMGALPDLDNODGFADVAVGAPLEDEHOGALTYLHG-TOSGVRPHPAORIAAAS 605
Db      518 GRFGAALTIVLGDVNEKILIDVAIGAPGEBENRGAVALPFGASBSGISPSHSQRITASQ 577
Qy      606 HALSYFGRSVDRDLDDGDLVDVAVGAGAAILLSSRPVHLTPSLVTPQALISVVO 665
Db      578 PRLQYFGQALSGGQDITDGLMDLAVGARGOYLLRLSLPVLKGVAMRSPPEVAKAVY 637
Qy      666 C-----RRROEAIVCLTALALCFQVTSRTPGRMDFQFMRFTASLIDETAARAPFDS 718
Db      638 CWEKPSALEADGATVCLTIQ-----KSSLDQIGDIQSSRPDLADPRLTSLRALFNET 692
Qy      719 GORLSRRLRLSVGNVTCQLHFNHVL-DSDYLRPAVALTVPALDNTTKPG-----PVAN 772
Db      693 KNPITLRKRTLGLG-IHCETILKLLPDCYEDVAVSPIILHNSILVAREPIPSQNLRPVLA 751
Qy      773 EGSPTSIOQLVPPSKDQCPDNCEVTDLVQLVNMDIRSKRAPVVRGGRKVLVSTLEEN 832
Db      753 VGSODLFTASLPFEKKNCGQDGLCEGD--LGVTLSFGGLQTL--LTVGSSLEINVIATVMN 806
Qy      833 KRENAYNTSLSTIFSNLHLASITPQRESPI--IVEC-AAPSAAHARL-----CYGHAV 883
Db      807 AGEDSYGTAVSLYYPAGLSHRRVSGAQKQPHQSALRLACETVFTEDGSLRSSCSVNHPI 866
Qy      884 FQTGAKVTFLEPEFSCSLISOVFGKLTASSDLSLRNGTLOENTQOTSATVIOYEPHLF 943
Db      867 FHEGSGTITVTFDVSRYKATTLGDRM-LMPSASSENKKAASSKATQLELPKIVAVYTHI 925
Qy      944 SSESTLHRYEVAHYGTLPVGPGEPEFKTTLRVONLGCYVVSGLIISALLPVAHAGNYFLS 1003
Db      926 SHQESSTKY--FNPAISDEKMKKEAHRVANNLSGRDLA-ISINWVVLVINGVAVM 982
Qy      1004 LSOVITNNASCIYONLTERPPRPVARELQHTNR--LNGSNTQCVAVRCHQLQAKGTE 1060
Db      983 VNEASQSILPCVSEH-----KPPHSDPLTQISRSFMLDCSIADICQFCADVPSPVQEE 1037
Qy      1061 VAVGLLRLVHNEFPRAKFKSLTVVSTFELGTEEGSVLQLTASRKSLEL-VYQTRRI 1119
Db      1038 LDFTLKGNLSFGVRETLOKKVAVSVAITPTDTSYSQLPGOEAMRQOMEMVLEDEEV 1097
Qy      1120 LISLWTLIGSVLGLLLALLVFCIMKLGF-AHKKIPEEKREE 1163
Db      1098 YNAIPIWSSVGCALLLLALITATLYKLGFFKRYHEMLEDKRED 1142

```

RESULT 12  
US-09-891-943-99

```

; Sequence 99, Application US/09891943
; Publication No. US20030077278A1
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vlieten, Monica
; TITLE OF INVENTION: No. US20030077278A1el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/891,943
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 99
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-891-943-99

Query Match      18.4%; Score 1126; DB 9; Length 1161;
Best Local Similarity 29.5%; Pred. No. 1,9e-82;
Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;

23 ENLDEHPRLLPFGPEAEFGYSVLQHVGGQRMMVLVGAPMDGPSGRRDYRCPVGGAH 82
   |||||
18 FLNDEEPTLIFQ-EDAGFGQSVOF--GSGR-LVVGAPLEAVANQTRGLVDC---AAA 70
   |||||
83 NARCAKHGLIDYQLGNSHPVNMHLMGMSLLETBDDGDFMCAPLMSACSSVSQSGIC 142
   |||||
71 TGMCPITPL-----HIREAVNMSLGLTLASTNGSLACGPTLHRVCGENSTSKSC 124
   |||||
143 ARVDASFOPOGSLAFTAOQCF-TYMDVIVILDGNSNI--YPMSEVQTFRLRLVGLFIDP 199
   |||||
125 LLLGSRWEIIQTVPDPATPECPHOEMDIVLIDGSSGSDIONDNFNOMKGFVQAVMGO--FEQ 182
Qy      200 EOIQVGLVOYGESPVHEMSLDGPFRTKEEVBAKKNLSREGETKTAQAINVACTEGESQ 259
Db      183 TDTLFLMQLNSLKHFTFOFRITSPSOOSLVDPIVQJLGL-TFTATGILLVVQTLFPH 241
Qy      260 SHGGRPEARLLVVTGDESH--DGEELPALKACAGATRYGIAVLGHYLRQRDPSSF 318
Db      241 KNGARSAKKILIVITDQKYPLEYSVDIIPAERKAGIIRYALGV-GHAF--QGPTA- 295
Qy      319 LREIRTIASDPDERFFENVTDAAALTDIDALGDRIFGLESHAENESSFGLMSQIGFS 378
Db      296 ROELNTISSAPPODHVFVDNFPALGSIQKQLOEKIYAVEGTQSRASSSFQHEMSQEGFS 355
Qy      379 THRLLKGLIFGNVAGAYDMGGSVLMLEGGHRLPPPRMALEDEFPALONHA-----AY 430
Db      356 TALTMGDLFLGAVGSFSW-----SGAFILYPRNMS-----PFIINMSQENVMDRSDY 402
Qy      431 LGYSVSMULRGGRRLFLSGAPPRHRKGVIAF-QLKQDGAVRVAOSLOGEIGSYFSGE 489
Db      403 LGYSTELALMKGVQNLVL-GAPRYOHTGKAVIFTQVSQW--RKAEVLTGTQISYFGAS 459
Qy      490 LCPIDTRDGTDTVLVLAAPMFLGPONKETGRVYVVLV--GQOSLLTLOGTQOPPEPD- 546
Db      460 LGSVDVSDGSTDLLILIGAPHYY--EOTRGQGVSCPLPRGRQVOMQCDAVLRGEQHPW 517
Qy      547 ARFGFAMGALPDLDNODGFADVAVGAPLEDEHOGALTYLHG-TOSGVRPHPAORIAAAS 603
Db      518 GRFGAALTIVLGDVNEKILIDVAIGAPGEBENRGAVALPFGASBSGISPSHSQRITASQ 575
Qy      606 HALSYFGRSVDRDLDDGDLVDVAVGAGAAILLSSRPVHLTPSLVTPQALISVVO 663
Db      578 PRLQYFGQALSGGQDITDGLMDLAVGARGOYLLRLSLPVLKGVAMRSPPEVAKAVY 635

```

QY 664 RDC-----RRGOEAVCLTAAICFOVTSRTGPRMDHOFYMRFTASLDEMTAGARAAD 716  
DB 636 YRCWEKPSALBAGDATTCTIO-----KSLDQDGIQSSVRFDALDPGRILTSRAIFN 690  
QY 717 GSGQRLSPRLRLISGVNTCEQLHFHVL-DSDYLAPVALTVTFALDNTTKPG-----PY 770  
DB 691 ETKNPFLTRKRTGLG-ICETLKLPLDCEVDVVSPIILHLNFSLVREPIPSQNLRPV 749  
QY 771 LNEGSPTSIOKLVPFSKDCGPDNECVTDVLQVNMIDIGSRKAPFVVRGRRVLVSTTL 830  
DB 750 LAVGSDDLFTASLPFEKNCGQDGLCEGD--LGVTLSFSGLOT--LTVGSSLELNIYIV 804  
QY 831 ENRKENAVNTSLISITFSRNLHLASLTTPQRES-----IKVEC-AAPSAHARL-----CSYGH 881  
DB 805 WNAEGSDYGTAVSLYYPAGLSHRVSGAQKQPHQSALRLACETVPTDEGLRSRCSVNH 864  
QY 882 PVFQTAKTFLLEFEFSGSSLSQVFGKLTASDSLERNGTLQENTAGTSAYIOYEPHL 941  
DB 865 PTFHESNGTFIVTFPVSYKATLGDRM-LMRASASSENKASSSKATFQLELPVKYAVYT 923  
QY 942 LFSESTLHRYEVHYPGTLVPVGPPEKTTLRVONIGCVVSGLIISALLPAVAHGNYF 1001  
DB 924 MISROESTKY--FNATSEKMKKEAHRVNNLSQDLA-ISINFWVPLLVNGVAVW 980  
QY 1002 LLSQVITNNASCIYQNTLTPPGRPVHPEELQHTNR--LNGSNTQCVVRCHLGQLAKG 1058  
DB 981 DVWMEAPQSPLCVSER-----KPOHSDPLTQISRSPLMDCSIADCLQPCDVPFSFVQ 1035  
QY 1059 TEVSUGLLRLVHNEFFRAKFKSLTVVSTFELGTEGVSUQLTEASRMSESLLE-VVQTR 1117  
DB 1036 BELDFTLKGNISFGWRETLOKKVLVVAELITFDTSVYSQLPQGEAFMRQMEWVLEED 1095  
QY 1118 PILISIMILIGSVLGLLALLLVFLMKLGF-AMKKIPEEKREE 1163  
DB 1096 EYVNAIPITMGSSVGALLLALITATLYKLGFKRYKEMLEDKPED 1142

RESULT 13  
US-09-350-259-99  
Sequence 99, Application US/09350259  
Patent No. US20020062008A1  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. US20020062008A1el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/350, 259  
CURRENT FILING DATE: 1999-07-08  
EARLIER APPLICATION NUMBER: 09/193, 043  
EARLIER FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173, 497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/266, 889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362, 652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943, 363  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 99  
LENGTH: 1161  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-350-259-99

Query Match 18.4%; Score 1126; DB 10; Length 1161;  
Best Local Similarity 29.5%; Pted. No. 1.9e-82;  
Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;

QY 23 ENLDENHRLRFRPRRAEFGYSVLQHVGGQKRMVLGAPWDBSSGRDRDVRCPVGGAH 82  
DB 18 FNDLVEEPTIFQ-EDAGFGQSVVDF--GQSR-LVVGAFLVVAANQTRLYDC--AAA 70

QY 83 NAPCAKHLGDYOLGNSHPAVNMHLGMSLETDDGCFMACAPLMSRACSSVPSGIC 142  
DB 71 TGMQCPILP-----HIRPEAVNMISGLTLASTNGSLNACGPTLHHVCGNSYSKSC 124  
QY 143 ARVDASFPQGSIAFPAQRCP-TYMDVIVLUDGNSI--YPMSEVOTLRRLVGLFTDP 199  
DB 125 LILGSRWEIITVPDAPTECPHOEMDIVFLIDGSSGIQDNFQNMKGCFVQAMGO--FEG 182  
QY 200 BQIQVGLQYGGSPPHWESLDDFRTKEEVAANKLRSREGETTAAIIVACREGSQ 259  
DB 183 TDTLALQVSNLKIHFTHFTQFTSPSQOSLVDPITVLKGL-TFTATGILTVTLQHH 241  
QY 260 SHGRPEARLIVVNTDESH-DGEBELPAALACAGVTRYGIAVLGHYLRORDPSSF 318  
DB 242 KMGARKSAKILIVITDQKXKDPLEYSYDVLPQAKAIGIRALGV-GHAF--QGPIA- 296  
QY 319 LREITIASDPERFEFVNTDEALITDVALGDRIFGLGSHABNESSFGLMSGIGFS 378  
DB 297 ROELNTISSAPQDHFVFDVNFALGSIQOLOEKIYAVEGTQSRASSSFQHEMSQEGFS 356  
QY 379 THRLLKDGILFGVGVAYMGSGSVLMLEGHRLPRPRMALEDEPRPALQNH-----AY 430  
DB 357 TALTYWDGFLAIVGSFSW-----SGGAPLYPPNMS-----PTFIMMSQENVDMRDSY 403  
QY 431 LGYSVSMILRGARLFLSGAPRFRHRGKVIAF-QLKXGAVRVVQSLQGEQIGSYFSGE 489  
DB 404 LGYSTELALMKGVQVNLV-GAPRYOHTGKAVIFITQVSNQW--RKAELVTGQIGSYFGAS 460  
QY 490 LCPDLTDDGTTDVLVLAAPMFLGPONKETGRVYVY-----LVGQSSLLTQGTLOPEBP 544  
DB 461 LCSVDVDSDDSTDLILIGAPPHY--EOTRGQVSCPLPRGRVQCCAVLRGE-QGHP- 516  
QY 545 QDARFGPMGALPDNODGFADVAAGAPLEPGHOGALVLYHG-TOSGVRPHPAORIAAS 603  
DB 517 -WGRFGAALTVLGDVNEKXLDVAILGAPGEOBNRCAVTLFHGASGSGISPSHSQRIASSQ 575  
QY 604 MPHALSYFGRSVDGLDDGDLVDVAVAGAAIILSSRPVILHLPSELVTPQAISVYQ 663  
DB 576 LSPRLQYFGQALSGGQDLTDQGLMDLAVGARGQVLLHSLPLVKGVAMRSPVEVAKAV 635  
QY 664 RDC-----RRGOEAVCLTAAICFOVTSRTGPRMDHOFYMRFTASLDEMTAGARAAD 716  
DB 636 YRCWEKPSALBAGDATTCTIO-----KSLDQDGIQSSVRFDALDPGRILTSRAIFN 690  
QY 717 GSGQRLSPRLRLISGVNTCEQLHFHVL-DSDYLAPVALTVTFALDNTTKPG-----PY 770  
DB 691 ETKNPFLTRKRTGLG-ICETLKLPLDCEVDVVSPIILHLNFSLVREPIPSQNLRPV 749  
QY 771 LNEGSPTSIOKLVPFSKDCGPDNECVTDVLQVNMIDIGSRKAPFVVRGRRVLVSTTL 830  
DB 750 LAVGSDDLFTASLPFEKNCGQDGLCEGD--LGVTLSFSGLOT--LTVGSSLELNIYIV 804  
QY 831 ENRKENAVNTSLISITFSRNLHLASLTTPQRES-----IKVEC-AAPSAHARL-----CSYGH 881  
DB 805 WNAEGSDYGTAVSLYYPAGLSHRVSGAQKQPHQSALRLACETVPTDEGLRSRCSVNH 864  
QY 882 PVFQTAKTFLLEFEFSGSSLSQVFGKLTASDSLERNGTLQENTAGTSAYIOYEPHL 941  
DB 865 PTFHESNGTFIVTFPVSYKATLGDRM-LMRASASSENKASSSKATFQLELPVKYAVYT 923  
QY 942 LFSESTLHRYEVHYPGTLVPVGPPEKTTLRVONIGCVVSGLIISALLPAVAHGNYF 1001  
DB 924 MISROESTKY--FNATSEKMKKEAHRVNNLSQDLA-ISINFWVPLLVNGVAVW 980  
QY 1002 LLSQVITNNASCIYQNTLTPPGRPVHPEELQHTNR--LNGSNTQCVVRCHLGQLAKG 1058  
DB 981 DVWMEAPQSPLCVSER-----KPOHSDPLTQISRSPLMDCSIADCLQPCDVPFSFVQ 1035  
QY 1059 TEVSUGLLRLVHNEFFRAKFKSLTVVSTFELGTEGVSUQLTEASRMSESLLE-VVQTR 1117  
DB 1036 BELDFTLKGNISFGWRETLOKKVLVVAELITFDTSVYSQLPQGEAFMRQMEWVLEED 1095

```
QY      1118 P L I L W L I G S V G L L L L L L V F C L M K L G F F - A H K I P E E K R E 1163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1096 E V Y N A I P I I M G S S V G A L L L L A L I T A L T Y L G C F F K H Y K E M L E D K P D 1142

RESULT 14
US-09-891-943-55
; Sequence 55, Application US/09891943
; Publication No. US20030077278A1
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. US20030077278A1 Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/891,943
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-891-943-55

Query Match      18.4%; Score 1124.5; DB 9; Length 1161;
Beet Local Similarity 30.1%; Pred. No. 2,5e-82;
Matches 358; Conservative 207; Mismatches 486; Indels 139; Gaps 45;

QY      20 C S P N L D E H P R L F P G P P A E F G Y S L Q H V G G O R M L V G A P W D G S G D R G D V Y R C P V G 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      17 C H G S N L D V E P I V F - R E D A S F G Q T V O F - - G G S R - L V G A P L E A V A V O T G R L Y D C - - - 69

QY      80 G A H N A P C A K G H L G D V Q L G N S S H P A V M H L G M S L L E T D G G C G M A C A P L M S R A C G S S V F S 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      70 - - - - A P - A T G M C Q P I V L - R S P L E A V M S L G L S L V A T N N A Q L L A G C P T A Q R A C V N M V A K 123

QY      140 G I C A R V A S F O P G S L A P T A O R C P T Y - M D V I V I L D G S N S I - - Y P M S E V O T F L R L V G K L F 196
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      124 G S C L L G S S L O F T I O A V P A S M P E C P R O E M I A F L I D S G S I N O R D P A Q M D F V K A L M G E - F 182

QY      197 I D P E O I O V G L V O Y G E S P V H E M S L G P R T K E E V V R A K N L S R R E G R E T K T A Q I M V A C T E G 256
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      183 A S T S T L - F S L M Q Y S N I L K H F T F E B K N I L D P Q S L V D P I V O L G L - T Y T A T G I R T V M E L 240

QY      257 F S Q S H G R E P A A R L L V V T D G E S H - D G E E L P A L A C E A G R V T R G I A V L G H Y L R R O R D P 315
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      241 F I S K S G S R S A K K I L V I T D G Q K Y R D P L E Y S D V I P A D A G A G I I R V A I G V D A F - - - O E P 296

QY      316 S S F L E I R T I A S D P D E R F F E N Y T D E A L D I D A L G D R I F G L E G S A E V E S S F G L E M S Q I 375
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      297 T A - L K E L N T T G A P P D H V F K V G N F A L S I O R Q L O E K I F A L E T O S R S S S F O H E M S O E 355

QY      376 G F S T R H L K D G I L F G W G A V D M G S V L M L E G H R L F P P - - - - R M A L E D E P P A L O N T A 428
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      356 G F S S A L T S D G P V L A N G S F S W - - - - - S G G A F L Y P P N T R P I F I M N S G E N - - - - V D M D 403

QY      429 A I L G S V S S M L R G G R L F L S G A P P R R H G K V I A P L K D G A V A Q S I O S Q I S Y F S 488
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      404 S Y L G S T A V A F W K G V S L L L - G A P R H Q H G K V I F - T O B A R H W R P K S E V R G Q I S Y F S A 461

QY      489 E L C P L D T D G T D V L L V A P M F L G P O N K E T G R V V Y V L V - G O O S L L T L O G T I O P E P P D - 546
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      462 S I C S V D V D D G S K D V L I G A P H Y - - E O T R G G O V S F P P V G R G R Q C A T I L H G E G H W 519

QY      547 A R F G P A M G A L P D L N D G F A D V A V G A P L E D H O G A L Y L Y H G - T O S G V R P H P A O R I A A S M P 605
```

```
Db      520 G R F G V A L T V L D D V N G D N L A D V A I G A P G E E S R G A V Y I F H G S A R L E I M P S P O R T V G S O L S 579

QY      606 H A L S Y F G R S V D G R L D L O G D D L V D V A V G A Q A I I L S R P I V H L P S L E V T F O A I S V V G R D 665
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      580 L R L O Y F G S I S G G D L T D D G L V D L A V G A Q G V L I R S I P L A K E I S I F A F M E V A K A Y O 639

QY      666 C R R R - - - - - G O E A V C L T A A L C F O Y T S R T G R M D H - O F Y M R F T A S L D E M T A G A R A A P D G 717
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      640 C M E R P T V L E A G E A T V C I T - - - - - V H K S P D L L G N V G S V R Y D A L D P G R L I S A I D E 693

QY      718 S G O R L S P R R L S V G N V T C E O L H F H V L D - T S D Y L R P V A L T V T F A L - D N T T P R - - G P V L N 772
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      694 T K N C T L T R K T L G L G D - H C E T V K L L P D C V E D A V S P I L R I N F S I V R S A S P R N L H P V I A 752

QY      773 E G S P T S I Q V L P F S D C G P D N E C V T D I V L Q V N M D I R S G R K A P F V R G G R R V I S T T L E N 832
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      753 V G S Q D H T A S L P F E K N C Q E L L C E D G I S T F N F S - - - - G L O V L V G S S P E L T V T V W N 807

QY      833 R K E N A V N T S L S I I F S R N L H L A S L T P O R E S - - - P I K V E C - A A P S A H A R L - - - - C S V G H P V F 884
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      808 E G E D S Y G L V K F Y P P A G S Y R R V T G T O Q P H Q Y P L R A C E A P A Q E D L R S S C S I N H P I F 867

QY      885 Q T G A K V T L L E P E F S C S L S Q V F E K L T A S S D S L E R N G T L O E N T A - Q T S A Y I O Y - - - - - 937
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      868 R E G A K T E M I T P D V S Y K A F L G D R L - - L I R A K A S S E N N P D N T K T A F O L E L P V K Y T V Y T L I 925

QY      938 - - - - - E P H L F S S - - - - - E S T L H R Y E V H Y G T L P V G P G E P F K T T L A V O N L G C V V A G L 985
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      926 S R O E D S T N H V N F S S H G R R O E A H R Y R V N N L S P L - - - - - K L A V N E M F V P V L L N G V 977

QY      986 I I - S A L L P A V A H G N Y F L S L S Q V I T N N A S C I V O N L T E P P G P V P H E E L O H T N R L N G S N T Q 1044
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      978 A V M D V T L S S P A Q G - - - - - - - V C S Q - - - - - M K R P Q A P D F L T O Q R S V L D C S I A D 1021

QY      1045 C Q V V R C H L G Q L A K T E V S G L L R L V H N E F F R A K F S I T V S T E L G T E B G S V L O L - - T E 1102
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1022 C L H F E C D I P S I D I O D E L F I R G N I S F G M V S Q T L O E K V L S E A E I T P D T S Y S O L P C O E 1081

QY      1103 A - - - - S R M S E S L E V Y Q T R P I L S I M L I L G S V L G I L L A L L V F C L M K L G F 1149
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1082 A F L R A Q V E T T L E B Y V E P I - - - - - F L V A G S S V G G L L L A L I T V V L Y K L G F 1127

RESULT 15
US-09-350-259-55
; Sequence 55, Application US/09350259
; Patent No. US20020062008A1
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. US20020062008A1 Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/350,259
; CURRENT FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/173,497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-350-259-55
```

Query Match 18.4%; Score 1124.5; DB 10; Length 1161;  
 Best Local Similarity 30.1%; Pred. No. 2.5e-82;  
 Matches 358; Conservative 207; Mismatches 486; Indels 139; Gaps 45;

QY 20 CSEFNIDEHHPRLFPPEPEAEFGYSVLQHVGGQRMVLGAPMDGSGRGDVRCPY 79  
 DB 17 CHGSNIDVEEPIVF-REDAASFGQTVVOF--GGSR-LVWGAPLEAAVAVQGTGLYDC-- 69  
 QY 80 GANAPACAGHLDYOLGNSHHPAVMMHIGMSLLETDDGGGFACALPMSACGSSVFS 139  
 DB 70 ---AP-ATGMCOPIYL-RSPLAVMNSIGSLVTNTNNAQLACGPTQORACVKNMYAK 123  
 QY 140 GLCARVDASFQGS LAPTAQRCPY-MDVVILDSNSI--YPMSEVOTFLRLVKGLE 196  
 DB 124 GSCILLGSSLQFIQAVPASMPECPRQEMDIAPLIDSGSINGQDFQMDPVYALMGE-F 182  
 QY 197 IDPEQIQVGLVOYGESPVHEWSLGDFTREEVVRAAKNLSRREGRETQAQAIMVACTG 256  
 DB 183 ASTSTL-FSLMOYSNLIKTHFTPEFKNILDPSLVDPIVOLQGL-TYTATGIRVMEEL 240  
 QY 257 FSGSHGRPEEARLLVVVTDGESH--DGEELPALKACEAGRVTRVGIIVLGHYLRQRDP 315  
 DB 241 FHSKNGSRSAKKILLVITDQKRPLEYSDVIRPADKAGIIRVAIGVDAF---QEP 296  
 QY 316 SSFLREIRTIASDPDERFFFNVTDEALTDIVDALGDRIFGLESHAENESSFGLMSQI 375  
 DB 297 TA-LKELNTIGSAPPDQHVFKVGNFALRSIQRLQEKIFALIGTQSRSSSFQHEMSQE 355  
 QY 376 GSTRHKLKQILFGMVGAYDWGGSVLMBSGHRLEPP-----RMALDEFPPLQNH 428  
 DB 356 GFSSALTSQPVLAGVGSFSW-----SGGAFLYPPNTRPTFINMSOEN-----VDMRD 403  
 QY 429 AYLYGVSSMLRGGRFLFSGAPRRHGRKVIAPQLKKQGAVRVAQSLQGEQISFGS 488  
 DB 404 SYLGISTAYAFMKGNHSLIL-GAPRHQHGKVIIF-TQEARHMRKSEVRGTQISYFGA 461  
 QY 489 ELCPLETDREDGTDVLLVAAPMELGPONKRTGRVYVLYV-GQOSLTLTQTLQPEPPD- 546  
 DB 462 SLCSVDVDRDGSXDLVLIGAPHYY--EQRGQGVSVFPPVGRVGRWQCEATLHGEGHPW 519  
 QY 547 ARPGFAMGALPDINOGFADVAGARLEBDHOGALYHNG-TOSGVRRHPAQRIAAAMP 605  
 DB 520 GRFGVALTVLGDVNGNLDVAIGARGEESRGAVYIFHGASRLIEIMPSPSORVTSQLS 579  
 QY 606 HALSYGRSVDGRLDDGDDLVVAAGACAAIILSSRPVHLTPSLVETPQALISVQGRD 665  
 DB 580 LRLQYFGQSLSGGQDLTQDGLVDLVAQGHVLLRSPLIKVELSIRAPMEVAKAVYQ 639  
 QY 666 CRRR-----GQEAVALTAALCFQVTSRTPGRMDH-QEYMRFTASLDEWTAGARAAPDG 717  
 DB 640 CWERTPTVLEAGATVCLT-----VHKSPDLGNVQGSVRYDLALDPRLISRALFDE 693  
 QY 718 SCGRSLSPRLRLSVGNVTCBQLHFVLD-TSDYLRVALTVTFAL-DNTTKP---GPVLN 772  
 DB 694 TNACTLTGRTKLTGLGD-HCEIVKLLPDCVEDAVSPILIRLNFSLVRDSASPRNLHPVLA 752  
 QY 773 EGSPTSIGLIVPSSKDCGPDNECVTDVLQVMMIDIGSRKAPVYVGRGRKVLVSTLLEN 832  
 DB 753 VGSQDHITASLPEPKCKOELLCEBDLGISFNFS-----GLOYLVGSGPELTVTYTVWN 807  
 QY 833 RENAVNTSLSIIFSRNLHLASLTPQRES---PIKVEC-AAPSAHARL-----CSVGHPV 884  
 DB 808 EGEDSYGLVVKFYYPAGLSYRRVYTGQGHQVYRLACEAPRAQEDLRSSGCSINHPIF 867  
 QY 885 QTGAKYTFLLBEFFSCSSLSLQVFGKLTSSDSLENGTLQENTA-QTSAVYQY----- 937  
 DB 868 REGAKTFMITPDVSYKAFLDRL--LLRAKASSENNKPDYNTAFQLELPVKYTYVYTLI 925  
 QY 938 -----EPHLFSS-----ESTLHRYEVHAPYGTLPVGPPEFKTLRAVQNLGCYVVSGL 985  
 DB 926 SROEDSTNHVNFSSSHGGRQEAHRYRYNNLSPL-----KLAVRVNFVVPVLLNGV 977  
 QY 986 II-SALLPAVAGNYFLSLQVITNNASCIVONLTPPGRPVHPELOHTNRLNGSNTQ 1044

DB 978 AWMVDTLSSPAQG-----VSCVSO--MKPQNPDDFLTQIRRSVLDCSIAD 1021  
 QY 1045 CQVACHGQOLAKGTEVSGLRLVHNEFFPRAKSLTVSTPELGTREBSVLQ--TE 1102  
 DB 1022 CLHFRCDIPSLDIODELDFILRGNLSPGWSQTLQEKVLVSEAEITPDTSVSQLPQE 1081  
 QY 1103 A---SRMSESLLEVYQTRPILISLWILIGSVLGLLLALVFCIMKLG 1149  
 DB 1082 AFLRAQVETTELFEYVVEPI-----FLVAGSSVGLLLALITVYLYKLG 1127

Search completed: July 16, 2003, 08:12:49  
 Job time : 89.5295 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:49:17 ; Search time 45.9578 Seconds  
(without alignments)  
747.132 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106  
Sequence: 1 MELPVTHTLPLVPLTGLC.....GFAFKKIPBEKREKLEQ 1167

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/ECTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1131.5	18.5	1161	1 US-08-173-497-2	Sequence 2, Appli
2	1131.5	18.5	1161	1 US-08-286-889-2	Sequence 2, Appli
3	1131.5	18.5	1161	1 US-08-485-618-2	Sequence 2, Appli
4	1131.5	18.5	1161	1 US-08-362-652-2	Sequence 2, Appli
5	1131.5	18.5	1161	2 US-08-605-672-2	Sequence 2, Appli
6	1131.5	18.5	1161	2 US-08-482-293A-2	Sequence 2, Appli
7	1131.5	18.5	1161	2 US-08-943-363-2	Sequence 2, Appli
8	1131.5	18.5	1161	4 US-09-193-043-2	Sequence 2, Appli
9	1131.5	18.5	1161	4 US-09-688-307A-2	Sequence 2, Appli
10	1126	18.4	1161	2 US-08-485-618-99	Sequence 99, Appli
11	1126	18.4	1161	2 US-08-605-672-99	Sequence 99, Appli
12	1126	18.4	1161	2 US-08-482-293A-99	Sequence 99, Appli
13	1126	18.4	1161	2 US-08-943-363-99	Sequence 99, Appli
14	1126	18.4	1161	4 US-09-193-043-99	Sequence 99, Appli
15	1126	18.4	1161	4 US-09-688-307A-99	Sequence 99, Appli
16	1124.5	18.4	1161	4 US-09-193-043-55	Sequence 55, Appli
17	1124.5	18.4	1161	4 US-09-688-307A-55	Sequence 55, Appli
18	1118.5	18.3	1161	1 US-08-485-618-55	Sequence 55, Appli
19	1118.5	18.3	1161	1 US-08-362-652-55	Sequence 55, Appli
20	1118.5	18.3	1161	2 US-08-605-672-55	Sequence 55, Appli
21	1118.5	18.3	1161	2 US-08-482-293A-55	Sequence 55, Appli
22	1118.5	18.3	1161	2 US-08-943-363-55	Sequence 55, Appli
23	1109	18.2	1161	1 US-08-485-618-53	Sequence 53, Appli
24	1109	18.2	1161	1 US-08-362-652-53	Sequence 53, Appli
25	1109	18.2	1161	2 US-08-605-672-53	Sequence 53, Appli
26	1109	18.2	1161	2 US-08-482-293A-53	Sequence 53, Appli
27	1109	18.2	1161	2 US-08-943-363-53	Sequence 53, Appli

28	1109	18.2	1161	4 US-09-193-043-53	Sequence 53, Appli
29	1109	18.2	1161	4 US-09-688-307A-53	Sequence 53, Appli
30	1108.5	18.2	1151	1 US-08-286-889-37	Sequence 37, Appli
31	1108.5	18.2	1151	1 US-08-485-618-37	Sequence 37, Appli
32	1108.5	18.2	1151	1 US-08-362-652-37	Sequence 37, Appli
33	1108.5	18.2	1151	2 US-08-605-672-37	Sequence 37, Appli
34	1108.5	18.2	1151	2 US-08-482-293A-37	Sequence 37, Appli
35	1108.5	18.2	1151	2 US-08-943-363-37	Sequence 37, Appli
36	1108.5	18.2	1151	4 US-09-193-043-37	Sequence 37, Appli
37	1108.5	18.2	1151	4 US-09-688-307A-37	Sequence 37, Appli
38	1105	18.1	1170	2 US-08-789-078-2	Sequence 2, Appli
39	1105	18.1	1170	2 US-08-752-633-2	Sequence 2, Appli
40	1105	18.1	1170	5 PCT-US95-04686-2	Sequence 2, Appli
41	1103	18.1	1155	1 US-08-286-889-46	Sequence 46, Appli
42	1103	18.1	1155	1 US-08-485-618-46	Sequence 46, Appli
43	1103	18.1	1155	1 US-08-362-652-46	Sequence 46, Appli
44	1103	18.1	1155	2 US-08-605-672-46	Sequence 46, Appli
45	1103	18.1	1155	2 US-08-482-293A-46	Sequence 46, Appli

## ALIGNMENTS

RESULT 1  
US-08-173-497-2  
; Sequence 2, Application US/08173497

; Patent No. 5437958  
; GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael  
APPLICANT: Van Der Vliet, Monica

TITLE OF INVENTION: No. 5437958el Human 2 Integrin Alpha  
TITLE OF INVENTION: Subunit

NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:

ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 S. Wacker Drive, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois

COUNTRY: USA  
ZIP: 60606-6402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/173,497  
FILING DATE:

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: No. 5437958and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/31363

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448

TELEX: 25-3856  
INFORMATION FOR SEQ. ID NO. 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-173-497-2

Query Match 18.5%; Score 1131.5; DB 1; Length 1161;  
Best Local Similarity 29.4%; Pred. No. 4.8e-100;  
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDHHRRLPGPEAFGYSVLQHYGGGRMLVGNPMGPGSDRGDVRGVGAH 82  
DB 17 FNLDVEPTIFQ-EDAGFGOSVQF-GGSR-LVVGAPLEVVANQGRLYDC---AAA 69

QY 83 NAPCAKHLGDYOLGNSHPAVNMHLGMSLLETDDGGFMACAPLMSACSSVFSGGIC 142  
 DB 70 TGMCGPIPL-----HIREPAVNMISGLTLTAASTNSRLLAGCPILHRCVGENSYKGGC 123  
 QY 143 ARVDASFPOGSLAPTAORCP-TYMDVIVLVDGNSI--YPMSEVOTFLRLVGLKFLIDP 199  
 DB 124 LLLGSRWEIITQVPDAPTECPHOEMDIVFLIDGSGSIDQNDPQMGKFGVQAVMGQ--PEG 181  
 QY 200 EOIQVGLVYGESPVHEMSLSDPRTKEVVRRAKLSRREGRTKTAQAIMVACTEGSFQ 259  
 DB 182 TDTLFLMOYSNLKIHFTFOPTSPSOGLVDPIVLKGL--TFTATGILTVVTLFPH 240  
 QY 260 SHGGRPEARARLVVVTDESH--DGEELPAALKACEAGRVTRYGIAVLGHYLRORDPSSF 318  
 DB 241 KNGARKSAKKILVITDQKYKPLEYSDVIPAOKAGIIRYALGV--GHAF---OGPTA- 295  
 QY 319 LREIRTIASDPDERFFPNVTDEALTIDVALGDRIFGLEGSHAENESSFGLMSQIGFS 378  
 DB 296 ROELNTISSAPPODHFKYDNFALGSIQKLOEKIYAVEGTQSRASSFOHEMSQEGFS 355  
 QY 379 THRLKXGILFGWYGAVDWGSVLMLEGGHLPPEPRALDEFPALQNA-----AY 430  
 DB 356 TALTMGGLFLGAVGSSFW-----SGGAFLYPPNMS-----PTINNSQENVMDRSDY 402  
 QY 431 LGYSVSSMLRGGRRLFLSGAPFRHRGKVIAP-OLKKGAVYVAOSLOGEIGSYFGSE 489  
 DB 403 LGYSTLALMKQONVL--GAPRYOHTGKAVIFTOYSRQK--RKKAEVGTQIGSYFGAS 459  
 QY 490 LCPDLTDROGTTDVLVAAPMELGPONKETGRVYVTV--GQGSLLTLOGTLQPEPPQD- 546  
 DB 460 LGSVDVSDGSDTDLILIGAPHYH--EOTRGGOVSVPPLPRGGRVQOCDAVIRGEGQHW 517  
 QY 547 ARGGFAMGLPDINOGFADVAAGAPLEDSHOGALVLYHG--TOSGRPHPAORITAAAMP 605  
 DB 518 GRGGAALTYLGDVNEBKLDIVATGAPGBOENRAVULFHASBSGSPSHSQIASQSQS 577  
 QY 606 HALSYFGRSVGRLDGDLDVDAVGAOGAAILLSRPIVHLTPSLLEVTPQAISVYQD 665  
 DB 578 PRIOYFGQALSGQDLTQDGLMDLAVAGAGQVLLLSLPLKVGAMRSPPEVAAYVR 637  
 QY 666 C-----RRRGQAVCLTALCFQVTSRTPEGRWDHQFYWRFTASIDETAGARAPFDS 718  
 DB 638 CWEKPSALEAGATVCLTIQ-----KSLDDQLGDIQSSVRPDLALPRLTSRAIFNET 692  
 QY 719 GQRLSPRLRLASGNTCEQLHFRVLD--TSDYLRPALVYTPALDNTKPG-----PVLN 772  
 DB 693 KNPFTLRRTKTLGIG-IHCEITLKLPLDCVEDVSPILHLNLSLVREPIPSPONLRPVLA 751  
 QY 773 EGSPTSIOQLVPEPSKDCGPDNECVTDVLQVNMIDINGSRKAPFVVGGRKXVLVSTTLEN 832  
 DB 752 VSGQDLFTASLPEPKKCGDGLCEGD--LGVTLISFGLQI--LTVGSSLELNVITVWN 806  
 QY 833 KRENAYNTLSLITFSNNHLASLTPORESP---IYEC-AAPSAAHRL---CSYGAFV 883  
 DB 807 AGEDSYGTAVSLYYPALSHRRVSGAOKOPHOSALACEVTEDEGRSSKCSVNHPI 866  
 QY 884 FQCGAVTFLLEPEFSCSLISQVFGKLTASSDLSLRNGTLOENTAQTSAVIOYEBHLF 943  
 DB 867 FHEGSGTITVITDVSFKATLGDRL--LMRASASSENKASSKATIQLEPKVAYVYMI 925  
 QY 944 SSBSTLHRYEHPYGTLPVGPPEFKTLRVONLGCYVVGSLIISALLPAVAGGVPLS 1003  
 DB 926 SNGEESTKY--FNFAISDEKKMEAEHRRYVNNLSGRDLA--ISINWVVLVINGVAVMW 982  
 QY 1004 LSGVITNNASCIYQNTTEPPGPPVPELOHTNR---LNGSNTQCVVRCHLQOLAKGTE 1060  
 DB 983 VNEAPQSLLPCVSEH-----KPOHSDFLTOISRSMLCSLADCLQCFCDVPSPFVOBE 1037  
 QY 1061 VASGLRLVHNEFFRAKFKSLTVASTFELTEGSGVLQTEASRWSESLTE--VVOGTRI 1119  
 DB 1038 LQFTLLGNLSPFGVNRETLQKXVLVVAVALITPTTSYSQIPGCEAFMRQMEKVLBEDEV 1097  
 QY 1120 LISLWILIGSVLGGILLALLLVFCWLKGF--AHKKIPPEEKRE 1163

DB 1098 YVAIPILMGSSVGLALLLALITATLYKLGFFKRYHKEMLEKPED 1142  
 RESULT 2  
 US-08-286-889-2  
 ; Sequence 2, Application US/08286889  
 ; Patent No. 5470953  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gallatin, W. Mich  
 ; APPLICANT: Van der Vieren, Monica  
 ; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 233 South Wacker Drive, 6300 Sear Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/286,889  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/173,497  
 ; FILING DATE: 23-DEC-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Williams Jr., Joseph A.  
 ; REGISTRATION NUMBER: P38,659  
 ; REFERENCE/DOCKET NUMBER: 27866/32168  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-474-6300  
 ; TELEFAX: 312-474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1161 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-286-889-2  
 Query Match 18.5%; Score 1131.5; DB 1; Length 1161;  
 Best Local Similarity 29.4%; Pred. No. 4,8e-100;  
 Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;  
 QY 23 FNLDHHRRLPPGPEAEFGYSVLOHVGCGORMLVGAPMWGSGDRRGDYRCVGAH 82  
 DB 17 FNLDHEERTITQ-EDAGFGQSVQF--GGR-LVVGAPLELVVANQGRVYDC---AAA 69  
 QY 83 NAPCAKHLGDYOLGNSHPAVNMHLGMSLLETDDGGFMACAPLMSACSSVFSGGIC 142  
 DB 70 TGMCGPIPL-----HIREPAVNMISGLTLTAASTNSRLLAGCPILHRCVGENSYKGGC 123  
 QY 143 ARVDASFPOGSLAPTAORCP-TYMDVIVLVDGNSI--YPMSEVOTFLRLVGLKFLIDP 199  
 DB 124 LLLGSRWEIITQVPDAPTECPHOEMDIVFLIDGSGSIDQNDPQMGKFGVQAVMGQ--PEG 181  
 QY 200 EOIQVGLVYGESPVHEMSLSDPRTKEVVRRAKLSRREGRTKTAQAIMVACTEGSFQ 259  
 DB 182 TDTLFLMOYSNLKIHFTFOPTSPSOGLVDPIVLKGL--TFTATGILTVVTLFPH 240  
 QY 260 SHGGRPEARARLVVVTDESH--DGEELPAALKACEAGRVTRYGIAVLGHYLRORDPSSF 318  
 DB 241 KNGARKSAKKILVITDQKYKPLEYSDVIPAOKAGIIRYALGV--GHAF---OGPTA- 295  
 QY 319 LREIRTIASDPDERFFPNVTDEALTIDVALGDRIFGLEGSHAENESSFGLMSQIGFS 378

```

296 ROELNTISSAPPODHVFNPAALGSIOKLOEKIVAVEGTQSRASSSFQHEMSQEGFS 355
379 THRDKGILFENGVAYDMGWSVLMLEGGHRLPPRMALEDEPPALONHA-----AY 430
356 TALTDGILFENGVAYDMGWSVLMLEGGHRLPPRMALEDEPPALONHA-----AY 402
431 LGYSVSMWLLRGGRRLFLSGAPRFRHGRKVIAT-OLKKGAVRVAOSLOGEOIGSYFSGE 489
403 LGYSTELAMKGVQNLVL-GAPRYOHTGKAVIFTOVSRQW--RKAAYVTGTIGSYFGAS 459
490 LCPRLDTRGRTDVLVAAPMFLGPNKGTGRVYVLV--GQOSLTLGTIGTQEPBPD- 546
460 LGSVVDSDGSDTLILIGAPHY--EOTRGQVSVCPRLRGQVQWQCDPAVLRGQHPW 517
547 ARFGFAMGALPDNODGFADVAVGAFLBEDHOGALVLYHG-TOSGVPRHAPORIAASWP 605
518 GRFGAALTYLGDVNEKIDVAIGAPGEQENGAVALFPHGASGSIQSPHSORIASQS 577
606 HALSYFGRSVDRDLDDGDLVAVAGAGAAILSSRPVHLTPSEVTPQAISVQWD 665
578 PRLQYFGQLSGGQDLTODGLMDLAVGAGVYLLRLSLVLKGVAMRFSPIEVAKAVYR 637
666 C-----RRGOEAVCLTALCFQVTSPTPGRMHQFMRTASIDENAGARAADGS 718
638 CMEKPSALEADATVCLTIO-----KSSLDOLGDIOSVRFDLADPRLTSRAIFNET 692
719 GGRLSRRLRSVGNVTCQQLHFNHVL--TSYLRPVALVTVALDNTKRG-----PVN 772
693 KNPILTRKRTLGIG-1HCETTLKLLPDCVEDVSPILHLNLSVREPISQONLRPVA 751
773 BGSPTSIOQLVPEFSKDCGPNECVNDVLQVNMDIRGSRKAPFVVGRRKVLVSTLEN 832
752 VGSQDLFTASLPEPKKCCGQDGLCEGD--LGVTLSFGQGT--LTVGSSLEINVLVTWN 806
833 KRENAYNTSLITIFSRNHLASITFORSP-----IKVEC-AAPSAHAR-----CSYGHV 883
807 AGEDSVGYVAVSYYPAGLSHRVSVGAOKOPHOSALRLACETVPTBEDGSRSSRCSVNHPI 866
884 FOTGAKVPTLPEFPEFCSSLLSQVFGKLFASSDSLEKNTLQENATQTSAYIOYEPHLF 943
867 FHEGSGITVTFEDSYKATLGDRLM-LMPASASSENKASSSQAIFOLEPVKAVYVYTI 925
944 SSESILHRYEVAHYGTLFVGPGEFETTLRLVONLGCYVVSGLISALPVAHAGNYFLS 1003
926 SGOESTKY--FNFAISDEKMKKEAHRVRVNNLSGRDLA-ISINFWVVLINGVAVMDV 982
1004 LSOVITNNASCIYONLTERRPPVPEPELOHTNR--LNGSNTQCCVAVRCHLGOLAKGTE 1060
983 VMEASQSILPCVSEH-----KPPHSDPFTOISRSPMLDCSIADCLQFCDVPSPSVQRE 1037
1061 VSVGLRLVHNEFFRPAKFKSLTVVSTFELGTEBEGVLOLTSASRSEBLE-VYQTRRI 1119
1038 LDFTLKGNISFGWVRETLQKVALVSVVAITPTDSYVSQLPGOEAPMRQOMEMVLEBDE 1097
1120 LISLWLLIGSVLGLLLALLVFLCMLKGLFF-AHKKIPEEKREE 1163
1098 VVAIPIMSSVYCALLLALLATLITATLYKGLFPRKHYKEMEDKPED 1142

```

```

STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: William J. Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-2
Query Match 18.5%; Score 1131.5; DB 1; Length 1161;
Beet Local Similarity 29.4%; Pred. No. 4.8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;
23 ENLDEHNRRLPPGEPEAFEGYSVLQHVGGGOMLVGAPMDPSGDRGDRYRCVGAH 82
17 FNLDVEEPTIQQ-EDAGSFGQSVOF--GGR-LVVGAPLELVANQGRLYDC--AAA 69
83 NAPCAKGLGYQLGNSHPVNNHLSLLETDDGPFMACAPMSAACSSVSSGIC 142
70 TGMCPRIPL-----HIREPAVNNSLGLTLASTNGSRLLACGPTLHRCGENSYSKSC 123
143 ARVDASFOQSLAPTAQRCPTWMDVIVLVDGNSI--YMSSEVQFLRLVLGKLTIDP 199
124 LLLGRMELIQTVPATATECHQEMDIYFLIDGSSIDQNFQMKGVQVAMQ--FEG 181
200 EQIQVLQVGYGSSPVHNSLDFRTKEEVNBAKRLSRREGETKYTAQAINVACTEGFSQ 259
182 TDTLPAALQVSNLKIHFPTQFRTSPQOSLVDEIVOLKGL-FETATGILTVTQLPHN 240
260 SHGGRPEAKRLVVTWDESH-DGEELPAALKACAGAVTYGIAVLGHYLRQDPSPF 318
241 KNGARKSAKILITVTDQKYKDPLEYSDVLPQAEKAGIIRYALGV-CHAF--QGPTA- 295
319 LREIRTIASDDPERFPVNTDEAALTDIVDALGDRIFGLGSHANESSFGLMSQIGFS 378
296 ROELNTISSAPPODHVFNPAALGSIOKLOEKIVAVEGTQSRASSSFQHEMSQEGFS 355
379 THRDKGILFENGVAYDMGWSVLMLEGGHRLPPRMALEDEPPALONHA-----AY 430
356 TALTDGILFENGVAYDMGWSVLMLEGGHRLPPRMALEDEPPALONHA-----AY 402
431 LGYSVSMWLLRGGRRLFLSGAPRFRHGRKVIAT-OLKKGAVRVAOSLOGEOIGSYFSGE 489
403 LGYSTELAMKGVQNLVL-GAPRYOHTGKAVIFTOVSRQW--RKAAYVTGTIGSYFGAS 459

```





```

Db 638 CMEKPSALAEADATVCLTIQ-----KSLDQDGIQSSVRDLDALDPRLTSRAIFNET 692
Qy 719 GORLSRRLRLSVGNVTCQQLHFHVL-D-TSDYLRPVALTVPALDNTTKPG-----PVLN 772
Db 693 KNPFLTRKRTLGIG-1HCETLKLRLPDCEVDVSPILHLNLSVREPIPSQNLRPVLA 751
Qy 773 EGSPTSIOQLVPPSKDCGPDNECVTDVLQVNMIDIRGSRKAPVVRGGRKVLVSTTLEN 832
Db 752 VGSODLFTASLPFEKKGCGDGLCEGD--LGVTLSFSGLOT--LTVGSSLEINVIIVTWN 806
Qy 833 RKENAVNTSLIIFSNLHLASLTPORES-1KVEC-AAPSAHARL-----CSVGHV 883
Db 807 AGEDSTIGTVASLYPAGLSHRVSGAQKQPHOSALRLACEVPTDEGRSSRCVNHPI 866
Qy 884 FOTGAKVFLLEFEFSCSSLLSOVGEKLTASSDSLENNGTLOENTAOQTSAYIOYPHLLF 943
Db 867 FHGSGNGFTIVFDVSYKATLGDRL-1MRASASSENKXSSSKATFQLELPKXAVYIMI 925
Qy 944 SSESILHRYEHPYGTLPVGPGEPEKFTLRYONLGCYVSGLIISALLPVAHGANFELS 1003
Db 926 SRQESTKY--FNFATSDEKMKKEAHRVANNLSQRLA-1SINFWVVLNGVAVMDV 982
Qy 1004 LSOVTNNAACVQNLTEPRGPVHPEELQHTNR---LNGSNTCCQVVRCHLQGLAKGE 1060
Db 983 VNEAPSQSLPCVSE----KPPQHSDFLTQISRSPLDCLADCLQECVDPSPSVQOE 1037
Qy 1061 VSVGLRLVHNEFPRAKESLTVVSTFELGTEEGSVQLTEASRWSSEILE-VVQTRPI 1119
Db 1038 LDFTLKNLSFGWVAVETLQKVLVVSVAEITFDTISYQLPGQEAFMRAQMVMLEDEV 1097
Qy 1130 LISLWILIGSVLGLLLALVFCWKLGF-AHKKIPBEKREE 1163
Db 1098 YNAIPIIMGSSVGLLLALITATLYKLGFFKRYKEMLEKPED 1142

RESULT 5
US-08-605-672-2
; Sequence 2, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Seair Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-605-672-2

Query Match      18.5%; Score 1131.5; DB 2; Length 1161;
Best Local Similarity 29.4%; Pred. No. 4.8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

Qy 23 FNLDHHRRLPPGPEAFEGSYVLOHVGSGRMVLVCAPMWPGSGDRGDYRCVGAH 82
Db 17 FNLDVEEPTIFQ-EDAGGFGSVQF--GGRS-LVWGAPLVAVANQGRLYDC---AAA 69
Qy 83 NAPCAKGIADYQJGNSHPAVNNHIGMSILETDODGGFMACAPLMSRACSSVPSGIC 142
Db 70 TGMCPPIPL-----HIREPAVMSLGLTLASTNGSLHLCGFLRHVCENSSKSC 123
Qy 143 ARVDASFQPGSLAPTAQCP-TYMDVYIVLDGNSI--YPMSEVQTEFLRLVGLFIDP 199
Db 124 LLGSRWIIQTVDPATPECHQEMDIYFLIDGSGSIDQNDFNQKGFVQAVMGQ--PEG 181
Qy 200 EIQVGLVQYESPHEMSLDFRTKEVVAAKLSRBERETKTAQAINVACIEGFSQ 259
Db 182 TDTLPAIMOYENLKHHTFQFRTPSPQSLVDPYOLKPL-FTATGILTVVQTLPH 240
Qy 260 SHGGRPEARLILVYVTDDESH-DEBELPALKACEAGRVTEYGIAVLGHYLRORDPSS 318
Db 241 KNGARKSKAKILYITDQOKYKDPLEISDVYPOAEKAGIIVYALGV-GHAF---OGPTA- 295
Qy 319 LREIRTTASDDEEFFNVNTDEALTIDVALGDRIFGLEGSHAENESSFGLMSQIGFS 378
Db 296 ROELNTISSAPPOQHVFVDNFFALGSIQKQLEKIVAEQTGRASSFQHEMSQEGFS 355
Qy 379 THRLKDGILFGWGAIVMGGSVLMLEGGHRLFPPRMALDEFPALONHA-----AY 430
Db 356 TALTMDGIEFLAVOSFSW-----SGGAFLYPPNMS-----PTFIMSOENVDMDRSY 402
Qy 431 LGYSVSMILRGGRRLFLSGAPRRHRGKXIAF-OLKKGDAVRYAQSLOGEOISYPOSE 489
Db 403 LGSTIELALMGVQNLV-LGAPRYOHTGKAVIFTOVSQW--RKAAEVTGTOIGSYFGAS 459
Qy 490 LCPIDTRDGTVDVLVAAPMFLGPONKETERVVVYLV--GQGSULTLQGTLOPEPPQD- 546
Db 460 LCSVDVDSGSDTDLILIGAPHY--EOTRGQGVSCPLPRQORVOMCCDALRGEQHPW 517
Qy 547 ARFGFAMGALPDLNODGFADVAVAPLEDHOGALVYHG-TQSGVRPHPAQRIAAAMP 605
Db 518 GRFGAALTVLQDVNEDKILDAVIGAPGEQNRGAIVYFHGASBSGISPSHQRIASSOLS 577
Qy 606 HALSYFGRSYDGRDLQDGDLDVDAVAGOGAAILLSSPIHLPLPSLEVTQQAISVQRD 665
Db 578 PRLQTFGQALSGGDLTODGLMDLAVGARGVTLRLSLPVTKVGAHKESEVEVAKAYR 637
Qy 666 C-----RRRGOAVCLTALACFOVTSRTPGRMDHOFYMFRTASLDEMTAGAAPDGS 718
Db 638 CMEKPSALAEADATVCLTIQ-----KSLDQDGIQSSVRFDLADGRITSRAIFNET 692
Qy 719 GORLSRRLRLSVGNVTCQQLHFHVL-D-TSDYLRPVALTVPALDNTTKPG-----PVLN 772
Db 693 KNPFLTRKRTLGIG-1HCETLKLRLPDCEVDVSPILHLNLSVREPIPSQNLRPVLA 751
Qy 773 EGSPTSIOQLVPPSKDCGPDNECVTDVLQVNMIDIRGSRKAPVVRGGRKVLVSTTLEN 832
Db 752 VGSODLFTASLPFEKKGCGDGLCEGD--LGVTLSFSGLOT--LTVGSSLEINVIIVTWN 806
Qy 833 RKENAVNTSLIIFSNLHLASLTPORES-1KVEC-AAPSAHARL-----CSVGHV 883

```

```

Db      807 AGEDSYGVTVSLYVAGLSHRVSGAQKQPHOSALALACETVTEDEGRSSRCVNHPI 866
Qy      884 FQTGAQVTVLLEFEFSCSSLLSQVFGKLTASSDLSERNGLQNTQTSAYIOYEPHLLF 943
Db      867 FHGSGNGTIVTFDVSYSKATLGDRLM-LMRASASSENKASSKATQLELPVYKAVYVTMI 925
Qy      944 SSESILHREYVHPYGLTPVGPGEFPTTLRVQNLGCVVSGLLISALLPAVAHGNYFLS 1003
Db      926 SROESTKY--FNFAISDEKMKAEHRVRVNNLSQRDLA-ISINFWVVLNGVAVMDV 982
Qy      1004 LSOVITNNASCIYQNLTEPPGPVHPEELQHTNR---LNGSNTQCQVACHLQOLAKTE 1060
Db      983 VNEAPQSILPCVSEK-----KPPQHSDFLTQISRSPMLDCSIADCLQFRCDVPSFSQEE 1037
Qy      1061 VSVGLLRLVHNEFRPAKRSKSLTVVSTFELGTEGSLQLTASRSSESLF-VYQTRPI 1119
Db      1038 LFTFLKGNLSFGVWRRTLOKVLVSVABITFTPTSVYSQLPQGEAFMRQMEVLEDEV 1097
Qy      1120 LISLWILIGSVLGLLLALVLVFCMLKLGFF-AHKKIPEEKREE 1163
Db      1098 YNAIPIIMGSSVGLALLLALITLTYKLGFPRKHVYEMLEDRPD 1142

```

## RESULT 6

US-08-482-293A-2  
Sequence 2, Application US/08482293A

Patent No. 5831029

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESS: 103

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Bortun

STREET: 233 South Wacker Drive, 6300 Sear Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/482,293A

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/173,497

FILING DATE: 23-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,889

FILING DATE: 5-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,652

FILING DATE: 21-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 37866/32684

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-482-293A-2

Query Match 18.5%; Score 1131.5; DB 2; Length 1161;  
Best Local Similarity 29.4%; Pred. No. 4,8e-100;  
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

```

Qy      23 FNLDEHPRLPPEPEAEFGYSVLQHVGGGRMVLGAPWDPGSDRRGDRYRCVGAH 82
Db      17 FNLDEHPRLPPEPEAEFGYSVLQHVGGGRMVLGAPWDPGSDRRGDRYRCVGAH 69
Qy      83 NARCAKHLGDIYQLGNSSHPANNHLMGSLLETDDGDFMCAPLMSACSSVSQIG 142
Db      70 TCMCPILP-----HIREAVNMSGLTLASTNGSRLLACGPTLHVCGENSYSKSC 123
Qy      143 ARVDSFPOGSLAFTAPQRC-tywDVIVLIDGNSI--YPMSEVQTLRLVLGTLFIDP 199
Db      124 LLLSRMELIQVPAPATECHQENDIVFLIDGSSIQNDPNQMKGVQAVMGQ--PEG 181
Qy      200 EGIQVLQVQGESPVHNSLDFPTKEEVNRAKNLSREGETKTAQAINVACTEFSQ 259
Db      182 TDTLPALQYSNLKIHFTQFRTPSQSLVDPIVLKGL-TFTATGILTVVTLFPH 240
Qy      260 SHGGRPEARLLVVTVDESH--DGEELPALKACAGVTVTGIAVLGHYLRQDRPSF 318
Db      241 KNGARKSAKILIVITDQKYKDPLEYSDVLPQAEKAGIRYALGV-CHAF--QGPYA- 295
Qy      319 LREIRTIASDPDERFFNVNTEAALTDIVDALGRIFGLEGSHANESSFGLMSQIGFS 378
Db      296 ROELNTISSAPRQDHVFVNDNFALGSIQKOLEKIVAVEGTQSASSSFQHEMSQEGFS 355
Qy      379 THRLKQILFGVAGAYDWGSLVLEGGHRLPPRMALEDEPPALQNH-----AY 430
Db      356 TALTMDGFLGAGVFSW-----SCGAFIYPPMS-----PFIMNSQENVDMDRSY 402
Qy      431 LGYSVSNLGGRRFLSLGAPRRHGRKYLAF-OLKKGAVRVAQSLQSGQISYFQSE 489
Db      403 LGYSTELALMGVQNLV-GAPRYOHTKAVIYFQVSQW--RKKAETGTQIGSYFAS 459
Qy      490 LCPDLTDRDGTDLVLAAPMFLGPONKETERVYVYLV--GQSSILTLTGTLQEPPOD- 546
Db      460 LCSVDVDDSGSTDLLIGAPHY--EQRGQVSVCP-PRQGRVOMQCDVLRGQGHFW 517
Qy      547 ARFGFAGALPDLDGFAVAVGAPLEDHOGALYLVHG--TQSGVRPHAPQRIAAAMP 605
Db      518 GRFGAALTVLGDVNEDKLIDVAGAPGEQNRGAVYLVHGHASGSSGISPSHQRIASOLS 577
Qy      606 HALSYFGRSVQGRDLDDDDLVDAVAGQAAILLSSPIYHLTSLTQALSVQORD 665
Db      578 PRLQYFGQALSGGQDLTDGDMDLAVGARQVLLRLSLPVKVGAMRFSPEVAKAYR 637
Qy      666 C-----RRRQEAVALTAALCFQVTSRTGRMDHQFYMRFSTASLDEMTAGARAAPDGS 718
Db      638 CMEERKSALEAGDAIVCLTIQ-----KSLDQDLGDSVAFDLDLGRILSRALFNET 692
Qy      719 GQRLSPRLRLSVGNVTCEQHLFHLVD--TSYLRPVALTLTPALDNTTKPG-----PVLN 772
Db      693 KNPILTTRKRTLGIG-ICETELKLLPDCVEDEVSPHIIHLNFSLVREPIPSQNLRPVLA 751
Qy      773 EGSFISIQKLVPEFSGDCGPDNECVTDVLYQNMIDRGSRKAPFVVRGGRKRVLTLEN 832
Db      752 VGSQDLFTASLPFEKNCQGDGLCEGD--LGVTLSFSGLQTL--LTVGSLLELVNLYVTWN 806
Qy      833 KENAYNTSLGIFSRNHLASLTPQRESF-----IKVEC-AAPSAHML-----CSGVHPV 883
Db      807 AGEDSYGVTVSLYVAGLSHRVSGAQKQPHOSALALACETVTEDEGRSSRCVNHPI 866
Qy      884 FQTGAQVTVLLEFEFSCSSLLSQVFGKLTASSDLSERNGLQNTQTSAYIOYEPHLLF 943
Db      867 FHGSGNGTIVTFDVSYSKATLGDRLM-LMRASASSENKASSKATQLELPVYKAVYVTMI 925
Qy      944 SSESILHREYVHPYGLTPVGPGEFPTTLRVQNLGCVVSGLLISALLPAVAHGNYFLS 1003
Db      926 SROESTKY--FNFAISDEKMKAEHRVRVNNLSQRDLA-ISINFWVVLNGVAVMDV 982

```

QY 1004 LSOVITNNASCIIVQNTLTPPPVHPEELQHTNR--LNGSNTQCVVRCHLGOLAKTE 1060  
DB 983 VMEASQSLPCVSE-----KPOHSDPLTOISRSPMLDCSIADCLQPCVPSFVGE 1037  
QY 1061 VSVGLRLVHNEFFRAKFKSLTVSTFLGTEEGSVQLTEASRSESLLE-VVQTRPI 1119  
DB 1038 LDFTLKGNLSFGWRETLOKVLVSVAEITFDTSVYSQLPGQEAEMRAQMEMVLEDEB 1097  
QY 1120 LISLWILIGSVLGGLLLLALVFCMLKLGFF-AHKKIPPEEKREE 1163  
DB 1098 YNAIPIIMGSSVGLALLLLALITATLYKLGFFKRYKEMLEDEKPED 1142

RESULT 7  
US-08-943-363-2  
Sequence 2, Application US/08943363  
Patent No. 5837478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seair Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-2

Query Match 18.5%; Score 1131.5; DB 2; Length 1161;  
Best Local Similarity 29.4%; Pred. No. 4.8e-100;  
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDEHRLPFGPPREAFEGYSVLQVGGORMLVGAPMDPSGDRGDRGVRCVPVGGH 82  
DB 17 FNLDEHRLPFGPPREAFEGYSVLQVGGORMLVGAPMDPSGDRGDRGVRCVPVGGH 82  
QY 83 NACPAGKHLGVDYQVQNGSSHPAVNMILGMSLLETTDGGGFMACAPLWSRAQSSVFSGGIC 142

DB 70 TGMCPPIPL-----HIREAVNNSLGLTLASTNGSRLLACGPTLHARVCGENSYSKSC 123  
QY 143 ARVDASFOPOGSLAPTAQCP--TWMDVIVILVDGSNSI--YPMSEVQFLRLVGLFLIDP 199  
DB 124 LLISGRMEIQTVPATTECHQEMDIYFLIDSSSIQNDPNQKGVQVAMQO--FEG 181  
QY 200 EOIQVGLVOYGEPSVHENSISDPFKTEVVRANKLSREGETKTAQAINVACTEGFSQ 259  
DB 182 TDTLPAHQVSNLKIHFTHFTQFRTPSQOSLVDPVQLKGL-TTATGILTLVTVQLPH 240  
QY 260 SHGGRPEARLLVYVTDSEH--DGEELPALKACAGAVTRYGLAVLGHYLRQRQDPSDF 318  
DB 241 KNGARKSAKILIVITDQKYKDPLEYSVDVIPAOKAGIIRYALGV--EHA--QGPTA- 295  
QY 319 LREIRTIASDPDERFFNVNTEDEALTDVDAIGRIFGLEGSHAENESSFGLMSQIGPS 378  
DB 236 ROELNTISSAPPQDHVFNDFNFAALGSIQKOLQETIVAVEGTQSRASSFQHEKQBGFS 355  
QY 379 THRLKDGILFGWAGYDWGSLWLEGGHRLFPFPMALDEFPALQNH-----AY 430  
DB 356 TALTWIDGLFLGAVGSSFW-----SGCAFILYPRMS-----PFIMNSQENVDMDSY 402  
QY 431 LGYSVSNMLGGRRLFLISGAPRFHNRKVIAT-OLKQDGAVRAVQSLQGEQISYFSGE 489  
DB 403 LGYSTELALMWGVQNLV--GAPRYOHTGKAVIFTOVSROW--RKKAETVGTQISYFGAS 459  
QY 490 LCPILDTDRGTTDVLVAAPMFLGPQNKETGRVYVYLV--GQOSLLTLOCTLOPPPD- 546  
DB 460 LCVVDVDSGSDTDLIGAPHY--EOTRGQVQSVCP--PRQGVQWQCDVAVRGQGHFW 517  
QY 547 ARFGFAMGALPDLNODGFADVAVGAFLBEGHAGALYLYHG--TOSGVRPHAPQRIAAAMP 605  
DB 518 GRFGALVIVLGVNBDKIDVAIGAPGEHNGAVYLFHGAESGSISSHGRISASQLS 577  
QY 606 HALSYFGSVNDGRLLDGDIDLVDVAVGAQAAILLSPPYVHLTPSLVTPQALSVQPD 665  
DB 578 PRLQYFGALSGGQDLDTDGMDLAVGAGQVLLRLSLPLVKVGAHMFSPVEVAKAYR 637  
QY 666 C-----RRRQGAVALCTALCFQVTSRTGRMHDQVYMFSTASLDMWTGARAFAFGS 718  
DB 638 CWEKPSALEAGDATVCLTIO-----KSLDQLGIGQSVFDDLDLGRLTSAIFMET 692  
QY 719 GQRLSPRLRLSVGNVNTCEQLHFVLD--TSDYLRVALTVTFALDNTTPG-----PVLN 772  
DB 693 KNPITLRKRTIGLG--IHETIKLLPDCVEDVVSIIILHNFSLVREPIPSQNRPLVA 751  
QY 773 EGSPTSIGKLVPSKDCGPDNECVTDVLQVNMIDIRGSKAPFVVRGRRKRVLSTILEN 832  
DB 752 VGSQDLFTASLPFEKNCCQDGLCEGD--LGVTLSFGHQOT--LTVGSSLEINVIIVTWN 806  
QY 833 RKENAVNYSLSIIFERNHLASLTPQRESF-----IKVBC-AAPSAHAF-----CSVGHV 883  
DB 807 AGEDSYGTVAVSLYYPAGSHRRVSGAQOPQOSARLACETVPFEDEGLRSRCSVNPI 866  
QY 884 FQTGAKVTFLEFEFSCSLISOVFGKLTASSDSILERNGTQENTQTSAYTQYEPHLLF 943  
DB 867 FHESNGNFIYTFDVSXATYATGDM--LMRAASSENNAASSSKATFOLELVKAVYIMI 925  
QY 944 SSESTLHRYEVHPYGTLPVGPPEFKTTLRVQNLGCVVVGIIISALLPAVAGGNFLS 1003  
DB 926 SROESTKY--FNFAISDEKMKKEAHRHYRVNNSISORDLA--ISINFWPVLINGVAVDV 982  
QY 1004 LSOVITNNASCIIVQNTLTPPPVHPEELQHTNR--LNGSNTQCVVRCHLGOLAKTE 1060  
DB 983 VMEASQSLPCVSE-----KPOHSDPLTOISRSPMLDCSIADCLQPCVPSFVGE 1037  
QY 1061 VSVGLRLVHNEFFRAKFKSLTVSTFLGTEEGSVQLTEASRSESLLE-VVQTRPI 1119  
DB 1038 LDFTLKGNLSFGWRETLOKVLVSVAEITFDTSVYSQLPGQEAEMRAQMEMVLEDEB 1097  
QY 1120 LISLWILIGSVLGGLLLLALVFCMLKLGFF-AHKKIPPEEKREE 1163  
DB 1098 YNAIPIIMGSSVGLALLLLALITATLYKLGFFKRYKEMLEDEKPED 1142

```
RESULT 8
US-09-193-043-2
; Sequence 2, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193, 043
; EARLIER FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/1173, 497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286, 889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362, 652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943, 363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-193-043-2

Query Match      18.5%; Score 1131.5; DB 4; Length 1161;
Best Local Similarity 29.4%; Pred. No. 4.8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY      23 FNLDEHHPRLPFGPEAPBEGYSVLQVHGQGMVLGAPMDPSGRGVVRCVPVGAH 82
DB      17 FNLDEEPTIFQ-EDAGGFGQSVQF--GGR-LVVGAPLEVVAHQGTGIDC---AAA 69
QY      83 NAPCAGHLGDYOLGNSHPAVNMHLMGSLLETDGCGFMACAPLMSRAGSSVFSGGIC 142
DB      70 TGMCGQIPL-----HIRPEAVMNSLGLTLAASTNSRLLAGCPILHRGGENSVKSGC 123
QY      143 ARVDASFQPGQSLAPTAQRCF-TYMDVIVVLGDSNSI--YPMSEVOTFLRLVGLKFLDP 199
DB      124 LLLGSMWEIQTVPDAPTPCEPHQEMDIVFLIDSGSIDONDPMQMGFYQAWVGQ--FEG 181
QY      200 EGIQVGLVOYGESPVHWSLGDRTKEVYRAKNSRREGRTKAQIMVACTEGFQ 259
DB      182 TDTLFLMOTSNLTKHFTPTQFRISPSQOSLVDPVQLKGL-TFTATGLTVVTLQFHH 240
QY      260 SHGGRPEARLLVVTVDGESH--DGEELPAALKACEAGRYVYGIAYLGHYLRORDPSPF 318
DB      241 KNGARSAKKILIVTDGQKYKPLEYSDVTPQAEKAGIIRVAGV--GHAF--QGPTA- 295
QY      319 LREIRITASDPDERFFFNVTDEALTDIYDALCDRIFGLEGSIAENESSFGLMSQIGFS 378
DB      296 ROELNTISSAPPODHFVKYDNFALGSIQKQLOEKIYAAVEGTOSRASSSFQHEMSQEGFS 355
QY      379 THRLLKGIIFGMAVDAWCGSVLWLGSHRLPPRAALDEPPALQNH-----AY 430
DB      356 TALTMDELGLGAVGSSW-----SGAFLYPPNMS-----FTINNQENVMDRSDY 402
QY      431 LGYSVSMILRGGRRLFLSGAPFRFRHKGVIAP-QLKQDGAIVVAOSLOGEIOISYFGSE 489
DB      403 LGYSTELAMKQGNQVLT--GAPRYQHTGKAVIFTOYSRQW--RKAEIVTGTQISYFGS 459
QY      490 LCPDLTDRTGTTDVLVAAPMFLGPONKETGRVYVTV--GQOSLLTLTGTTQPEPPD- 546
DB      460 LCSVDVDSGSDTLITIGABHY--EOTRGQGVSVCPPLRGQVVOQCDAVLRGEGHPW- 517
QY      547 ARPGFMGALPDINOGFADVAVGAPLEGHOGALVLYHG--TOSGRPHPAQRIAAAMP 605
DB      518 GRFGAALTVLGVNDEKLIIDVAILGAPGEBENRGAVYLFHGASGSISSHQRIASSQUS 577
```

```
QY      606 HALSYFGRSVDRLDDDDLVDAVGAOAILLSSRPVHLTPSLVTPQALISVQORD 665
DB      578 PRLQYFGQLSGGQDQLTQDGLMDLAVGARQVLLRLSPVLKVGVAFFSPVEVAKAYVR 637
QY      666 C-----RRQGEAVCITLALCFQVTSRTPEQRMHQYMFMTASLDWETACARAAPQGS 718
DB      638 CWEKPSALEADATVCLTLIQ-----KSLDQLOGIQSSVREDLALDQRLTSRAIFMET 692
QY      719 GQRLSPRLRLRSVGNVTCQEOHLHFVLD--TSPLYRVALTVTFALDNTTKPG-----PVLN 772
DB      693 KNPILTRKRTLGLG--IHETLKLPLDCEVEDVSVIILHNFSLVREEPIPSQNRPLVA 751
QY      773 EGSPTSIOKLVPFSKDCGPDNECVTDLVLQVNMDIRSGRKAPFVVVGRGRRKVLSTLEN 832
DB      752 VQSQDLFTASLPFEKNGCQDGLCEGD--LGVTLSFGHQF--LTVGSSLEINVIIVTVWN 806
QY      833 KENMYNYSLSIIFERNHLASLTQRSP-----IKVVC-AAPSNHAL-----CSVGHV 883
DB      807 AGEDSYGTIVVSLYYPAGSHRRVSGAQKQPHQSALRLACETVPTEDEGLRSSRCSVNHPI 866
QY      884 FQTGAKVFLLEFEFSCSSLQVFGKLTASDSLERNGTQENOTSAVIOYEPHLF 943
DB      867 FHGSGNGFTIYTFDVSYATTLGDRM-LMRASSENNAASSSKATFQLELPVKYAVYMI 925
QY      944 SSESTLHRYEVAHPYCTLVPGGPEPEKTLTARVQNLGCVVYSGLLISALLPAVAGNYFLS 1003
DB      926 SRQESTYK--FNFATSDKKWEAEHRYRVNNSIQRDLA-ISINFVPLVNLGVAVMV 982
QY      1004 LSQVITNNAACLVNLTPEPBPVHPELQHTNR---LNGSNTQCVVCHGLQAKTE 1060
DB      983 VNEAPSQSLPCVSE-----KPOHSDFLTOISRPMIDCSIADLQRCVPSFSVQEE 1037
QY      1061 VSVGLRLRVHNEFFRAKFKSLTVVSTFELGTEGSSVQLTEASWSSESLLE-VYQTRPI 1119
DB      1038 LDTFLKMLSCQWAEETLOKVLVVSVAEITFDTSVYQLQGEFMFAQMEMVLEBDEV 1097
QY      1120 LISLWILIGSVLGLLALLLVFCLWKLGF-ARRKPIEEKREE 1163
DB      1098 VYAIRIIMSSVYGALLLALITATLYKLGPFKRYKEMLEBKPED 1142

RESULT 9
US-09-688-307A-2
; Sequence 2, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688, 307A
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193, 043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605, 672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173, 497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286, 889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362, 652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943, 363
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-688-307A-2

Query Match      18.5%; Score 1131.5; DB 4; Length 1161;
```



```

QY 143 ARVDASFOGSLAPTAORCP--TYMDVIVILDGNSI--YPMSEVOTFLRLVGLFTDP 199
DB 125 LILGSRWEIITQVDPATPECPHOEMDIIVFLIDGSGSIDQNDPFNGKGFVQAVMGO--FEG 182
QY 200 EOIQVGLVQGESPVHMSLGDRTKEEVRAAKNLSREGRETKTAQAIIVACTEGESQ 259
DB 183 TDTLFALMOYSNLKHIFFTQFTSPSQSLVDPVQLKGL--FTATGILTVVTLFPH 241
QY 260 SHGGRPEARALLVIVTDGESH--DGEELPALKACAGAVTRYGIAVLGHYLRQRDPSS 318
DB 242 KNGARKSAKILIVITDQKYPLEYSVDIIPAERKAGIRYAIQV--GHAFF--OGPTA 296
QY 319 LREIRTIASDPDERFFFNVTDEALTDIYDALGDRIFGLESHAENESSFGLEMSQIGFS 378
DB 297 RQELNTISSAPQDHFVKVNFALGSIQKOLEKIYAVEGTQSRASSSFQHEMSQEGFS 356
QY 379 THRUKGILFGWGVAVDMGGSVLMLEGGHRLFPFPAALDEFPALQNH-----AY 430
DB 357 TALTMGDLFLGAVGSSFSM-----SGGAFLYPPNMS-----PTFINMSQENVMDRDSY 403
QY 431 LGYSVSSMLRGGRRFLGAPRRFRHKGVIAP--OLKXQAVYVAOSLQEOIGSYFSGE 489
DB 404 LGISTLALMKGVQNLVL--GAPRYQHTGKAVITQVSROW--RKAEVITGTOIGSIFGAS 460
QY 490 LCPDLTDROGTTDVLVAAPMLGPNKKTGRVYVY-----LVGQSLTLTQGLQPEPP 544
DB 461 LGSVDVDSQSTDLILIGAPHYV--EOTRGQVSVCPPLPRGRVQWCDAILRGE--QGHF 516
QY 545 QDARFPGMAGALPDLNODGFADYAVGAPLEDHGQALYLYHG--TQSGVAPHPQRILAAAS 603
DB 517 -WGRFPAALTVLGDVNEKRLIDVAIGAPGEENRGAVYLFHGASGSGISPSHSQRILASSQ 575
QY 604 MPHALSYFGRSVGRJLDLGDGDLVDYAVGAQGAAILLSRPYIHLPPSLVFPQALSVQ 663
DB 576 LSRRLQYFQALSGSGDGLTODGLMDLAVGARQVLLRSLPYUKVAVNARFSEVEAKAV 635
QY 664 RDC-----RRRQEAVALTALCFQVTSRTFGRMDHOFYMRFTASLDEMTAGARAAPD 716
DB 636 YRCWEKPSALEAGDATVCLTIQ-----KSLDLQDLQISSVAFDIALDPGLTSRAIFN 690
QY 717 GSGQRISPRRLSVGNVTCQOLHFIIVLD--TSDYLAPVALTYVFALDNTTKRP-----PV 770
DB 691 ETKNPTLTKRLTGLD--IHCEITKLLPCVCEDEVASPIILHLNFSLVREPIBPQNLRYV 749
QY 771 LNEGSPSTIQKLVFPFGSKDCGPDNECVTDVIAQVNMIDIRGSRKAPFVVRGGRKVLSTTL 830
DB 750 LAVGSGDGLTASLPFRKNCQDGLCEGD--LGVTLSPSGIQT---LTVGSSLELVNIYIV 804
QY 831 ENRKENAVYTSLSIIFSRLHLASLTPQRESP---IKVEC--AAPSAAHRL---CSVGH 881
DB 805 WNAAGEDSYGVVSVLYYPAGLSHRVSGAQKQPHQSALRLACETVPTDEGLRSGSCSVNH 864
QY 882 PVFQGTAKTFLLEFFPFGCSLSLSQVFKULTASDSLENGTQIQTNTAQTSAVIOEPLH 941
DB 865 PTFHESNGSTFIYTFVVSYKATLGDRLM--LMRASASSENNKASSSKATFDELEPVKXAVYT 923
QY 942 LFSSESTLARVEYHPVGTLPVGPPEFKTTLRVONLGCYVVSGLIISALLPAVAGHGNF 1001
DB 924 MTSRQESSTKY--FNATSDKMKRAEHRVNNLSQSDLA--ISINFWPVULINVAW 980
QY 1002 LSLQVITNNAACIVONLTERPGRPVHPELOHTNR--LNGSNTQCVVRCGLQGLAKG 1058
DB 981 DVVMEAPSOQLPCVSEB-----KPPHSDFLTQISRSPMLDGSIADECLFRCDVPFSVQ 1035
QY 1059 TEYSVULLLVNHEFRRAKFKSLTVSVTFELCTEGSVLQULTEARMSBSLIE--VQNR 1117
DB 1036 BEIDFTLKGNLSFGWRETLQKKVLVSAEITFDTSVYSQLPQGAFAKRAQEMVLEBD 1095
QY 1118 PILISWILIGSVLGLLLALLLVFCLMKLGGF--AHKXIPEEKREE 1163
DB 1096 EYVNAIPITMGSSVGLLLALLTATLTYKLGFFKRYKEMLEBDKPED 1142

```

```

RESULT 11
US-08-605-672-99
Sequence 99, Application US/08605672
Patent No. 5817515
GENERAL INFORMATION:
APPLICANT: Galatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,672
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO.: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-605-672-99
Query Match 18.4%; Score 1126; DB 2; Length 1161;
Best Local Similarity 29.5%; Pred. No. 1,7e-99;
Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;
QY 23 FNLDEHNRLLPPGPEAFGYSLQHVGGGQRMVLVGAPWDGSPGDRGRDYRCFVGAN 82
DB 18 FNLVDEEPTIRQ--EDAGFGQSVQV--GGR--LVVGAFLVAVANQGRLYDC---AAA 70
QY 83 NARCAKGLGDYQLOGNSHPAVNNHLGMSLLETQDGGFMACAPLMSACSSSVSSGIC 142
DB 71 TGMCPRIPL-----HIREAVNMSLGLTLASTNGSRLACGFLHVCGENSYSKSC 124
QY 143 ARVDASFOGSLAPTAORCP--TYMDVIVILDGNSI--YPMSEVOTFLRLVGLFTDP 199
DB 125 LILGSRWEIITQVDPATPECPHOEMDIIVFLIDGSGSIDQNDPFNGKGFVQAVMGO--FEG 182
QY 200 EOIQVGLVQGESPVHMSLGDRTKEEVRAAKNLSREGRETKTAQAIIVACTEGESQ 259
DB 183 TDTLFALMOYSNLKHIFFTQFTSPSQSLVDPVQLKGL--FTATGILTVVTLFPH 241
QY 260 SHGGRPEARALLVIVTDGESH--DGEELPALKACAGAVTRYGIAVLGHYLRQRDPSS 318
DB 242 KNGARKSAKILIVITDQKYPLEYSVDIIPAERKAGIRYAIQV--GHAFF--OGPTA 296

```





```

Db 404 LGYSTELALMKGVQNLV-LGAPRYOHTGKAVIFTOVSROW--RKKAENVTOIGSYFGAS 460
Qy 490 LCPDPTDGTDTTDLVLAAPMFLGPONKENGRYVYV-----LVGOOSLLTLOGTLQPEPP 544
Db 461 LCSVDVSDGSDTDLILIGAPHY--EOTRGQVSVCPPLRGVOWCCDAVLKGE--QGHF 516
Qy 545 ODARFGFMGALPDLNODGFADYAVGAPLEDGHOGALVLYHG--TOSGVRPHPAORIAAAS 603
Db 517 -WGRFGAALTVLGDVNEKULIDVAIGAPGEBENKAVYLPHGASESGISPSHSQRIASSQ 575
Qy 604 MPHALSYPGRSVDRGLDGDLDLVDAVAGAGAAIILSSRPVHLTPSLVETPOAISVYQ 663
Db 576 LSPRLQYFCQALSGGODLTQDGLMDLAVAGRGVLLRLSPVLKGVAVNRFSEVAKAV 635
Qy 664 ROC-----RRRGQAVCLTAALCGQVTSRTGGRMDHOFYRFTASLEMTAGARAAD 716
Db 636 YRCWEKPSALEAGDATTCLITD-----KSSLDQGLDIOSSVAFDIALPGRILTSALIN 690
Qy 717 GSGORLSPRLRLSVGNVTCCEOLHFFVLD-TSDYLRPVALTVTALDNTTKRG-----PV 770
Db 691 ETKNPILTRKTLGLG-INCETLKLPLPCVEDVVSPIILHLNFSLVREPIPSQNLARV 749
Qy 771 LMBGSPISQKLVPEKDCGPDNECVTDLVLYQNMDIRSKRAPFVVRGGRKRVLSTTL 830
Db 750 LAVGSODLFTASLPFEKNGGODLCEGD--LGVTLSFSGIQT--LTVGSSLELVNIVTV 804
Qy 831 ENRKENVATSLIIFSRMLHLASLTPORESF-----IKVEC-AAPSAHARL-----CSVGH 881
Db 805 WNAAGEDSYGTIVSVLYYPAGLSHRVSGAQKOPHQSLRLACEVTPFEDEGLSSRCSVNH 864
Qy 882 PVFOGTAKVTFLEPFSSCSLLSQVFGKLTASSDSLEBNGTLOENTAOISAVIOYEPL 941
Db 865 PIFHEGNGTFIVTFVSVYKATLGDHM-LMRASASENNKASSKATFOLELPVKYAVYT 923
Qy 942 LSSSESTLRYEHPHRTGLPVGEGPEFKTLRLQNGICVYVSGSLISALLPAVAGKXNF 1001
Db 924 MTSROESTKY--FNATSDDEKMKKEAHRVYNNLSQBDLA-ISINFVWPVLNGLVAVW 980
Qy 1002 LLSQVITNNAACIVGNLTETPPGPVPHPELQHTNR--LNGSNTOQOVVRCHLGQLAG 1058
Db 981 DVVMEAPSSQLPCVSR-----KRPQHSDELTOISHPMLDCSIADCLFRCVPSFSQ 1035
Qy 1059 TEVSVGLRLVHNEFFRAKFKSLTVVSTFELCTEGSVLQLTLEASRWSESLLE--VVQTR 1117
Db 1036 BELDFTLKNLSFGWVRETLOKKVLVVAEITFDTSVVSQPLQCGAFRAQEMVLEED 1095
Qy 1118 PLIISIMILIGSVLGLLILALLVFLMKLGF-AHKTIPEEKREE 1163
Db 1096 EYVNAIPITMGSSVGLLILALITATLYKLGFEKRYKEMLEBKPBD 1142

```

## RESULT 13

```

US-08-943-363-99
Sequence 99, Application US/08943363
Patent No. 5837478
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5837478 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-99

Query Match 18.4%; Score 1126; DB 2; Length 1161;
Best Local Similarity 29.5%; Pred. No. 1,7e-99;
Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;

Qy 23 FNLDHHRRLFPPEAFEGFYSVLQHVGGGORMMLVGAQMPDGGDRGDVYRCVGAH 82
Db 18 FNLDVEEFTIFQ-EDAGFGQSVQF--GCSR-LVGAHPLEVVANQGRLYDC---AAA 70
Qy 83 NAPCAKGLIDGYOLGNSHPAVNMHLGNSLLETGDDGGFMACAPLMSFACSSVFSGIC 142
Db 71 TGMCPRLP-----HIREAVNMNSLGLTLAASSTGSRLLACGPTLHRVCGENSYSKSC 124
Qy 143 ARVDASFQPGSLAPTAORCP-TYMDVITYVLDDGNSI--YMWSEVQTLRLRLVGLFLDP 199
Db 125 LLLGSRWIIQTVDPATPECPHOEMDIYFLLDGSGSIDQDNFQMKGFVQAVMQQ--FEG 182
Qy 200 EIQVGLVQGESPVHEWSLDFRTKEEVRBAKNLSREGEKFTQAQIMWACTEGFSQ 259
Db 183 TDTLFALQVNSLNKIHFTTFQFRISPSQSLVDPYVQLKL-FTTALGILTVTQLFHN 241
Qy 260 SHGGRPEARLLLVVTDGESH-DGEELPALKACEAGVTRYGIAVLGHYLRQRPSSF 318
Db 242 KNGARKSAKKILIVITDQKYKDPLEYSDVIPQAEKAGIIRVAIGV-GHAF---QGPYA- 296
Qy 319 LREITIASDPERFFNVTDGALTDIVDLGRIPLBESHAENESSFGLMSQIGFS 378
Db 297 ROELNTISSAPPQOHVFVDNFPALGSIQKLOEKIVAVEQTQSASSSFHEMSQEGFS 356
Qy 379 THRLEDGILFGWVAGYDWGSGSVLMLEGGHRLFPPEMALEDEFPALQNA-----AY 430
Db 357 TALTYDGLFLGAVSGFSF-----SGGAFLYPPMS-----PFIIMSGEENDMRSY 403
Qy 431 LGYSVSMILRGRRRLFGAPRFRRHNGKVIAF-QLKKDGAVRVAQSLQGEQIGSYGSE 489
Db 404 LGYSTELALMKGVQNLV-LGAPRYOHTGKAVIFTOVSROW--RKKAENVTOIGSYFGAS 460
Qy 490 LCPDPTDGTDTTDLVLAAPMFLGPONKENGRYVYV-----LVGOOSLLTLOGTLQPEPP 544
Db 461 LCSVDVSDGSDTDLILIGAPHY--EOTRGQVSVCPPLRGVOWCCDAVLKGE--QGHF 516
Qy 545 ODARFGFMGALPDLNODGFADYAVGAPLEDGHOGALVLYHG--TOSGVRPHPAORIAAAS 603
Db 517 -WGRFGAALTVLGDVNEKULIDVAIGAPGEBENKAVYLPHGASESGISPSHSQRIASSQ 575
Qy 604 MPHALSYPGRSVDRGLDGDLDLVDAVAGAGAAIILSSRPVHLTPSLVETPOAISVYQ 663

```



```

QY      1118 PILISMLTIGSVLGGLLALLLALVFCMLKGF-  

      1096 EYVNAIPIIMGSSVGAALLLALITATLVYKLGFFKRYKEMLEDPED 1142
      DB

RESULT 15
US-09-688-307A-99
Sequence 99, Application US/09688307A
Patent No. 6432404
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6432404el Human Beta-2
FILE REFERENCE: 27866/36646
CURRENT APPLICATION NUMBER: US/09/688,307A
CURRENT FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 09/193,043
PRIOR FILING DATE: 1998-11-16
PRIOR APPLICATION NUMBER: 08/605,672
PRIOR FILING DATE: 1996-02-22
PRIOR APPLICATION NUMBER: 08/173,497
PRIOR FILING DATE: 1993-12-23
PRIOR APPLICATION NUMBER: 08/286,889
PRIOR FILING DATE: 1994-08-05
PRIOR APPLICATION NUMBER: 08/362,652
PRIOR FILING DATE: 1994-12-21
PRIOR APPLICATION NUMBER: 08/943,363
PRIOR FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 99
LENGTH: 1161
TYPE: PRT
ORGANISM: Homo sapiens
US-09-688-307A-99

Query Match      18.4%; Score 1126; DB 4; Length 1161;
Best Local Similarity 29.5%; Pred. No. 1.7e-99;
Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;

QY      23 FNLDEHPRLPFGPPPAEFGYSLQHVGGQRMVLVGAEMDGGSGRRDVRCPYGAH 82
      18 FNLDEEPTIFQ-EDAGCGGOSVQF--GGSRLVVGADPLEVVAANQTRLXDC--AAA 70
      DB

QY      83 NRPCAKGHGADVQLGSSHPAVVMHLMGSLLETDGCGGFMACAPLMSRAGSSVSSGIC 142
      71 TGMCOPIPL-----HIREBAVNMISGLTLASTNGSRLLACGPTLHRVCGENSISKSGC 124
      DB

QY      143 ARVDASFQPGSLAPTAQRCP--TYMDVIVLDSGNSI--YPMSEVQTFRLRLVGLKFLDP 199
      125 LLLGSFWEIITQVPRDAPTECPHOEMDIVFLIDSSGSDQNDPQMGKFOAVMGC--FEG 182
      DB

QY      200 EGIQVGLVOYGESPVHEWMLGDFRTKEEVVRAAKNLSRREGRETQAQIMVACTEGFSQ 259
      183 TDTLFLMQYSNLKLIHFTQFRTPSQOSLVDPIVQLKGL-TFTATGILTVVITQLFHH 241
      DB

QY      260 SHGGREARALVWYTDGSSH--DGEELPAAKACEAGRVRGIANVLGHYLRORPSSF 318
      242 KNGARSAKAILIVITDQKYPDLEYSVITPOAEKAGIIRYATIGV-GHAF--OGPTA- 296
      DB

QY      319 LREIRTIASDPDERFFPNTDEALTDIYDALGDRIFGLESHAENESSFGLEMSQIFPS 378
      297 RQELNITISSAPQODHFXTDNFALGSIQKOLEKIYAVEGTQSRASSFGHEMSQEGFS 356
      DB

QY      379 THRLLKDGILFGVAVDYGWGSVLMLEGGHRLFPPRMALDEPPPALQNTA-----AY 430
      357 TALTMGDLFLGAVGSSFSW-----SGAFLVPPNMS-----PTFINMSQENVMDRDSY 403
      DB

QY      431 LGYSVSMMLRGGRRLFLSGAPFRHRGVIAF-QLKDGAVRVASLOGEOIGSFGE 489
      404 LGYSTELAMKGVQNLVL-GAPRYOHTGRAVIFTQVSRQW--RKAAEVITGTOIGSYFGAS 460
      DB

```

```

QY      490 LCPDLTDRDGTDTVLVLAAPMFLGPONKETGRVYVY-----LVGOQSLLTLOGLQPEPP 544
      461 LCSVVDSDGSDTDLIGAPHY--EQTRGGQVSVCPJPRGRVQWQCDVAVLRG- QGHP- 516
      DB

QY      545 QDARFGFAMGALPDINOGFADVAVGAPLEBHGQALYLHYG-TOSGVRPHPAQIIAAS 603
      517 -WGRFGAALTVYGDVNEDEKLDVAILGARGEQENRGAAYLFFHGASGSI:SPHSORIASQ 575
      DB

QY      604 MPHALSVEGRSVDRGLDDGDLDVDAVGAQAAIILSSRPVHLTPSLLEVTPQALISVYQ 663
      576 LSPRLQYTGQALSGQDILTQDGLMDLAVGARGQVILLLSLVLKGVAMRSPVAVAAV 635
      DB

QY      664 RDC-----RRGGEAVCLTAALCFQVTSRTPGRMHQFYMRFTASLDEMTAGARAAD 716
      636 YRCWEKXSALBAGATVCLTIQ-----KSLDQGLQISSVRFPLALDPGLTSRALFN 690
      DB

QY      717 GSGQRLSPRLRLSVGNVTCEQLHNYID-TSDYLRPALVTPALDNTTBG-----PV 770
      691 ETKNPTLFRRTKLIGG-THCETLKLPLPCVEDVVSPIILHNFSLVREPIPSQNLBPV 749
      DB

QY      771 LNEGSPTSIQKLVPPSKDCGPDNECVTDVLVQVNMIDIRGSRKAPFVVGGRKVLVSTTL 830
      750 LAVSQDLFTASLPPEKNGQDGLCEBD-LGVTLSEGLQT--LTVGSSLELNVITV 804
      DB

QY      831 ENRKENAVNTSLISIFSRNLHLASLTTPQRESP---IKVEC-AAPSAAHRL---CSYGH 881
      805 WNAGEDSYGTVSLYYPAGLSHRVYSGAQKQPHQSALACETVTEDEGLRSSRCSYNH 864
      DB

QY      882 PVFQTAQVTFLEEFESCSLSQVFGKLTASSDSLERNGTLOENTQTSAYIOYEBHL 941
      865 PIFHEGNSGTPIVTFDVSKATLGDGM-LMRASASSENNKASSSKATFOLELPVYAAVT 923
      DB

QY      942 LFSSESTLHRYEVHPHYGTLPVGPGEPEFTTLRVQNLGCVVSGLITSLALPAVAHGYNF 1001
      924 MTSRQEBSTKY--RNFATSDKMKKEAHRVRVNNLSQDLA-ISTINWVPLVNGVAVW 980
      DB

QY      1002 LLSQVITNNASCIVQNLTEPPGPVPEPELOHTNR---LNGSNTQCCVAVCHLQOLAKG 1058
      981 DVVMEAPQSJPCVSEK-----KPPQHDPLTQISRSBMLDCSIADCLQFRCDVPSFSVQ 1035
      DB

QY      1059 TEVSGILRLVHNEFRRAKESLTVSTFPELGEBSVLDLUTERSKESILE-VVQTR 1117
      1036 BELDFTLKGNISFGVARETLQKXVLVSVASAITDTSVYSQLPQGEAFMRQOMENVLEED 1095
      DB

QY      1118 PILISMLTIGSVLGGLLALLLALVFCMLKGF-  

      1096 EYVNAIPIIMGSSVGAALLLALITATLVYKLGFFKRYKEMLEDPED 1142
      DB

```

Search completed: July 16, 2003, 07:57:46  
 Job time : 50.9578 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:52 ; Search time 10.1646 Seconds

(without alignments)  
1872.646 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025

Sequence: 1 GICARVDASFQSGSLAPFA.....FLREIRITASDPDERFFNNV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	473	46.1	272	A55348	integrin alpha-1 -
2	467.5	45.6	1151	A45226	integrin alpha-1 c
3	465.5	45.4	1180	A35854	integrin alpha-1 c
4	454	44.3	1178	S44142	VLA-2 protein homo
5	451	44.0	1170	I45914	integrin alpha 2 s
6	441	43.0	1181	A33998	integrin alpha-2 c
7	336	32.8	191	I47230	VLA-2 protein - pi
8	227.5	22.2	1747	A45974	collagen alpha 1(X
9	227.5	22.2	1857	S31212	collagen alpha 1(X
10	227.5	22.2	1888	S78476	collagen alpha 1(X
11	227.5	22.2	3124	A40020	collagen alpha 1(X
12	214	20.9	493	A33809	cartilage matrix p
13	205	20.0	496	A37979	cartilage matrix p
14	204.5	20.0	929	I51027	type XII collagen
15	201.5	19.7	337	A37797	collagen alpha 3(V
16	199	19.4	3051	S42373	hypothetical prote
17	194.5	19.0	500	S66522	cartilage matrix p
18	177.5	17.3	1179	A53213	integrin alpha-E c
19	176.5	17.2	3176	1 CGHJ3A	collagen alpha 3(V
20	175	17.1	724	A48569	antigen Em100 - E1
21	170.5	16.6	843	A40970	undulin 1 - human
22	169	16.5	741	T46488	hypothetical prote
23	158.5	15.5	1153	1 RWHU1B	cell surface glyco
24	158	15.4	1163	1 RWHU1B	lymphocyte functio
25	157.5	15.4	1153	2 S00551	leukocyte surface
26	156	15.2	550	2 T23760	hypothetical prote
27	156	15.2	2944	2 A54849	collagen alpha 1(V
28	154	15.0	2813	1 VWHU	von Willebrand fac
29	153.5	15.0	1163	1 RWHU1C	cell surface glyco

30	152	14.8	1170	2 S03308	cell surface glyco
31	142.5	13.9	712	2 A45638	immunodominant m
32	141	13.8	341	2 T32949	hypothetical prote
33	136.5	13.3	427	2 G00039	von Willebrand fac
34	129.5	12.6	414	2 P50323	von Willebrand fac
35	123	12.0	1286	2 A88396	protein M01E10.2
36	116.5	11.4	13055	2 T16580	hypothetical prote
37	114.5	11.2	567	2 T28797	hypothetical prote
38	112.5	11.0	661	2 T16597	hypothetical prote
39	110	10.7	643	2 T19549	hypothetical prote
40	107	10.4	449	2 T33257	hypothetical prote
41	103.5	10.1	2098	2 T18397	protein CTRP - mal
42	100	9.8	1029	1 S21369	collagen alpha 2(V
43	97.5	9.5	780	2 A34102	von Willebrand fac
44	95.5	9.3	1019	1 A32856	collagen alpha 1(V
45	95	9.3	445	2 B40970	undulin 2 - human

## ALIGNMENTS

RESULT 1  
A55348  
integrin alpha-1 - chicken (fragment)  
C/Species: Gallus gallus (chicken)  
C/Date: 10-Feb-1995 #sequence\_revision 10-Feb-1995 #text\_change 17-Mar-1999  
C/Accession: A55348  
R/Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.E.  
J. Biol. Chem. 269, 22811-22816, 1994  
A/Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1  
A/Reference number: A55348; MUID:94357930; PMID:7521332  
A/Accession: A55348  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-272 <KER>  
A/Cross-References: GB:U10114  
C/Suprafamily: unassigned collagens; von Willebrand factor type A repeat homology  
F/55-230/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match  
Best Local Similarity 46.1%; Score 473; DB 2; Length 272;  
Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;

Qy	1	GICARVDASFQSGSLAPFAQRCPTVMVIVLDGNSIYPWSEVQFLRLVGLFIDP	60
Db	30	GVCSNVSTFERNKAVASVQCKTQDIVLDGNSIYFRESVTALNSLRMDIGP	89
Qy	61	EQIQGLVQVGSPPVHNSLDGFTKEEVRAAKNLSRREGRETKTAQAINVACTEFSQ	120
Db	90	OOTQVGIQVQGVVHFVLYNTSTTEVMDAALRIRQGGTQWTALGIDTAREAEATTE	149
Qy	121	SHGGRPEARLLVYVTDSDGDELPAALAKCEGRTRGAINLVGYLRQRDPSSFL	180
Db	150	AHGARGVQKAVVITDGDSDHNVRLQEVLDKCEDENIQRPAIALLGSYSGNLSTKFKV	209
Qy	181	REIRITASDPDERFFNNV	198
Db	210	EEIKLSKPTKEKHFNNV	227

RESULT 2  
A45226  
integrin alpha-1 chain - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2000  
C/Accession: A45226  
J. Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.  
J. Biol. Chem. 268, 2989-2996, 1993  
A/Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.  
A/Reference number: A45226; MUID:93155124; PMID:8428973  
A/Accession: A45226  
A/Status: preliminary; not compared with conceptual translation  
A/Molecule type: mRNA









Db 309 PDSTHYVNV 317

RESULT 11

A40020

Collagen alpha 1(XII) chain precursor - chicken

N:Alternate names: fibrochimerin

C:Species: Gallus gallus (chicken)

C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 19-Jan-2001

A:Accession: A40020; A34485; B34485; A28037; S23014; S22254; S28811

R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shimomura, T.; Tanaka, H.; Nishida, Y.; Obayashi, Y. Cell Biol. 115, 209-221, 1991

A:Title: The complete primary structure of type XII collagen shows a chimeric molecule with a region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.

A:Reference number: A40020; MUID:92011662; PMID:1918137

A:Accession: A40020

A:Molecule type: mRNA

A:Residues: 1-3124 <YAM>

A:Cross-references: GB:D00824; NID:9222810; PIDN:BAA00701.1; PID:9222811

A:Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and, R:Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.

J: Biol. Chem. 264, 19772-19778, 1989

A:Title: Type XII collagen: A large multidomain molecule with partial homology to type I

A:Reference number: A34485; MUID:90062079; PMID:2584192

A:Accession: A34485

A:Molecule type: mRNA

A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>

A:Cross-references: EMBL:505137; NID:g211284; PIDN:AAA4635.1; PID:g211285

A:Accession: B34485

A:Molecule type: protein

A:Residues: 2772-2792; 2846-2873 <GOR2>

R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.

Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987

A:Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA c

A:Reference number: A28037; MUID:87317590; PMID:3476925

A:Accession: A28037

A:Molecule type: mRNA

A:Residues: 2960-2976, 'F', 2978-3074, 'AG' <GOR3>

A:Cross-references: EMBL:M17275; NID:g211649; PIDN:AAA4718.1; PID:g211650

A:Note: This sequence has been revised in reference A34485

R:Koch, M.; Bernasconi, C.; Chiquet, M.

Eur. J. Biochem. 207, 847-856, 1992

A:Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of

A:Reference number: S23814; MUID:92362621; PMID:1323460

A:Accession: S23814

A:Molecule type: protein

A:Residues: 'X', 1333, 'Q', 1335-1347, 1914-1928, 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517 <R>

R:Dublet, B.; van der Rest, M.

J: Biol. Chem. 262, 17724-17727, 1987

A:Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-

A:Reference number: S22254; MUID:88087065; PMID:3121603

A:Accession: S22254

A:Molecule type: protein

A:Residues: 2831-2832, 'T', 2834, 'R', 2836-2843; 3002-3014 <DUB>

R:Trub, J.; Trub, B.

Biochim. Biophys. Acta 1171, 97-98, 1992

A:Title: The two splice variants of collagen XII share a common 5' end.

A:Reference number: S28811; MUID:93042014; PMID:1420368

A:Accession: S28811

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>

A:Cross-references: EMBL:X67327

C:Genetics:

A:introns: 2845/3; 2863/3; 2867/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1

C:Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von Willebrand factor type I repeat homology; cell binding; coiled coil; connective tissue; disulfid

F:1-24/Domain: alternative splicing; cell binding; coiled coil; connective tissue; disulfid

F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>

F:24, 1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted

F:24-114/Domain: I11A #status predicted <I11A>

F:24-105/Domain: fibronectin type III repeat homology <FN3A>

F:137-301/Domain: von Willebrand factor type A repeat homology <VWA1>

F:332-425/Domain: I11B #status predicted <I11B>

F:332-414/Domain: fibronectin type III repeat homology <FN3B>

F:437-601/Domain: von Willebrand factor type A repeat homology <VWA2>

F:629-1178/Domain: I11C #status predicted <I11C>

F:630-711/Domain: fibronectin type III repeat homology <FN3C>

F:721-802/Domain: fibronectin type III repeat homology <FN3D>

F:812-895/Domain: fibronectin type III repeat homology <FN3E>

F:905-986/Domain: fibronectin type III repeat homology <FN3F>

F:1086-1169/Domain: fibronectin type III repeat homology <FN3G>

F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>

F:1384-2295/Domain: I11D #status predicted <I11D>

F:1384-1465/Domain: fibronectin type III repeat homology <FN3I>

F:1474-1557/Domain: fibronectin type III repeat homology <FN3J>

F:1566-1647/Domain: fibronectin type III repeat homology <FN3K>

F:1655-1738/Domain: fibronectin type III repeat homology <FN3L>

F:1756-1838/Domain: fibronectin type III repeat homology <FN3M>

F:1847-1928/Domain: fibronectin type III repeat homology <FN3N>

F:1937-2019/Domain: fibronectin type III repeat homology <FN3O>

F:2028-2110/Domain: fibronectin type III repeat homology <FN3P>

F:2119-2199/Domain: fibronectin type III repeat homology <FN3Q>

F:2207-2294/Domain: fibronectin type III repeat homology <FN3R>

F:2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>

F:2438-2440/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted

F:2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted

F:2571-2902/Domain: collagenous COL2 #status predicted <COL2>

F:2889-2891/Region: cell attachment (R-G-D) motif

F:2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>

F:2946-3048/Domain: collagenous COL1 #status predicted <COL1>

F:3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>

F:32, 1006, 1032, 1044, 1512, 1761, 2210, 2273, 2532, 2683/Binding site: carbohydrate (Asn) (cova

F:2780, 2789, 2836, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) #stat

Query Match 22.2% Score 227.5; DB 1; Length 3124;

Best Local Similarity 34.1%; Pred. No. 3.3e-12;

Matches 59; Conservative 32; Mismatches 67; Indels 15; Gaps 5;

QY 28 DVIVLDSNSI--YPMGEVOTFLRLVGLKFIPEQIQVGVGVGSPVHMSLGDPT 85

Db 1199 DIVLVDSNSIGRPNFTVNFISRIYEVDPDKQIGLAQSGPRTFEMINAVRT 1258

QY 86 KEEVRAKNLSRREGRETKTAQIMVACTGFSQSHGGRPEARLVLVVTGDSHDEE 145

Db 1259 KEALLDVATNLPYK-GGNTLTGMALDFLTKNFQGEALRRARXIGVLITDGSGQDDV 1317

QY 146 LPALKACEAGRVTRYGIATVGHYLRORDPSSFLRETRITASDPDEFRRNV 198

Db 1318 TPRSRLDEG--VELYIAIGI-----KNADEN---ELKQIATDPDDIHAYNV 1358

RESULT 12

A33809

Cartilage matrix protein precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 09-Mar-1990 #sequence revision 09-Mar-1990 #text change 21-Jul-2000

A:Accession: A33809; A26364

R:Kiss, I.; Deak, F.; Holloway Jr., R.G.; Delius, H.; Medhurst, K.A.; Frimberger, E.; Argr

J: Biol. Chem. 264, 8126-8134, 1989

A:Title: Structure of the gene for cartilage matrix protein, a modular protein of the ex

A:Reference number: A33809; MUID:8925546; PMID:2542265

A:Accession: A33809

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-493 <KIS>

A:Cross-references: GB:X12346; GB:X12347; GB:X12348; GB:X12349; GB:X12350; GB:X12351; GB

R:Argyraz, W.S.; Deak, F.; Sparks, K.J.; Kiss, I.; Goettlinc, P.F.

Proc. Natl. Acad. Sci. U.S.A. 84, 464-468, 1987

A:Title: Structural features of cartilage matrix protein deduced from cDNA.

A:Reference number: A26364; MUID:87092429; PMID:3025875

A:Accession: A26364

A:Molecule type: mRNA





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:42 ; Search time 5.1519 Seconds  
(without alignments)  
1594.036 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025  
Sequence: 1 GICARVDASFOQSGIAPTA.....FLREIRITASDPDERFFENV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1025	100.0	1167	1 ITAG_HUMAN	O75578 homo sapien
2	544	53.1	1189	1 ITAH_HUMAN	O90KX5 homo sapien
3	473	46.1	1285	1 ITAI_CHICK	O90615 gallus gall
4	467.5	45.6	1151	1 ITAI_HUMAN	P56199 homo sapien
5	465.5	45.4	1180	1 ITAI_RAT	P18614 rattus norv
6	454	44.3	1178	1 ITA2_MOUSE	O62469 mus musculu
7	451	44.0	1170	1 ITA2_BOVIN	P53710 bos taurus
8	441	43.0	1181	1 ITA2_HUMAN	P17301 homo sapien
9	227.5	22.2	1888	1 CAIE_CHICK	P13944 gallus gall
10	227.5	22.2	3124	1 CAIC_CHICK	P13944 gallus gall
11	214	20.9	493	1 CAMA_CHICK	P05099 gallus gall
12	213.5	20.8	3119	1 CAIC_MOUSE	O60847 mus musculu
13	209.5	20.4	3063	1 CAIC_HUMAN	O99715 homo sapien
14	205	20.0	496	1 CAMA_HUMAN	P21941 homo sapien
15	204.5	20.0	929	1 CAIC_NOTVI	O91145 nocophthalm
16	201.5	19.7	3137	1 CA36_CHICK	P15898 gallus gall
17	199	19.4	3767	1 MUA3_MOUSE	P34576 caenorhabdi
18	196	19.1	956	1 MTN2_MOUSE	O08746 mus musculu
19	194.5	19.0	500	1 CAMA_MOUSE	P51942 mus musculu
20	191	18.6	956	1 MTN2_HUMAN	O00339 homo sapien
21	186	18.1	624	1 MTN4_MOUSE	O08929 mus musculu
22	185.5	18.1	619	1 MTN4_HUMAN	O95460 homo sapien
23	177.5	17.3	1179	1 ITAE_HUMAN	P38570 homo sapien
24	176.5	17.2	3176	1 CA36_HUMAN	P12111 homo sapien
25	169.5	16.5	2482	1 VWF_PIG	O28833 sus scrofa
26	164	16.0	2813	1 VWF_CANFA	O28295 canis fami
27	160	15.6	1162	1 ITAD_HUMAN	Q13349 homo sapien
28	158.5	15.5	1152	1 ITAD_HUMAN	Q13349 homo sapien
29	158	15.4	1163	1 ITAL_MOUSE	P24063 mus musculu
30	157.5	15.4	1153	1 ITAM_MOUSE	P05555 mus musculu
31	156	15.2	2944	1 ITA7_HUMAN	O02188 homo sapien
32	154.5	15.1	1167	1 ITA7_MOUSE	O06077 mus musculu
33	154	15.0	2813	1 VWF_HUMAN	P04275 homo sapien

34	153.5	15.0	1163	1 ITAX_HUMAN	P20702 homo sapien
35	152	14.8	1170	1 ITAL_HUMAN	P20701 homo sapien
36	147	14.3	486	1 MTN3_HUMAN	O15232 homo sapien
37	143	14.0	452	1 MTN3_CHICK	O42401 gallus gall
38	140	13.7	547	1 COCH_CHICK	O42163 gallus gall
39	135	13.2	481	1 MTN3_MOUSE	O35701 mus musculu
40	120.5	11.8	550	1 COCH_HUMAN	O43405 homo sapien
41	120.5	11.8	552	1 COCH_MOUSE	O62507 mus musculu
42	110.5	10.8	639	1 CA1C_RABIT	O28902 corycolagus
43	100	9.8	1029	1 CA26_MOUSE	O02788 mus musculu
44	95.5	9.3	1019	1 CA16_CHICK	P20785 gallus gall
45	92	9.0	1018	1 CA26_HUMAN	P12110 homo sapien

## ALIGNMENTS

```

RESULT 1
ID ITAG_HUMAN STANDARD; PRT; 1167 AA.
AC O75578; OSUNH28;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-UN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-10 precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular chondrocytes;
RX MEDLINE=98352078; PubMed=9685391;
RA Camper L., Hellman U., Lundgren-Akerlund E.;
RT "Isolation, cloning, and sequence analysis of the integrin subunit
RT alpha10, a beta1-associated collagen binding integrin expressed on
RT chondrocytes."
RL J. Biol. Chem. 273:20383-20389(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart, and Endothelial cells;
RX MEDLINE=20169197; PubMed=10702680;
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,
RA Wang S.-X., Langley R., Kristensen G.W.;
RT "The integrin alpha10 subunit: expression pattern, partial gene
RT structure, and chromosomal localization."
RL Cytogenet. Cell Genet. 87:238-244(1999).
CC - FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC - SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10
CC ASSOCIATES WITH BETA-1.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN
CC MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.
CC - DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC - SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC - SIMILARITY: CONTRAINS 1 VWFA DOMAIN.
CC - SIMILARITY: CONTRAINS 7 FG-GAP REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; AF074015; AAC31952.1; -
DR EMBL; AF112345; AAF21944.1; -
DR EMBL; AF172723; AAF61638.1; -
DR HSSP; P17301; IAOX.
DR Genew; HGNC:6135; ITGA10.

```

```

DR MIM; 604042; -.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR Pfam; PF01839; FG-GAP 5.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS00234; VWF_A; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium; Magnesium.
FT CHAIN 1 22
FT SIGNAL 1 22
FT DOMAIN 23 1167
FT TRANSMEM 1123 1145
FT DOMAIN 1146 1167
FT REPEAT 38 97
FT REPEAT 97 97
FT REPEAT 167 350
FT REPEAT 365 427
FT REPEAT 428 482
FT REPEAT 483 545
FT REPEAT 546 605
FT REPEAT 608 660
FT DOMAIN 1134 1140
FT CA_BIND 494 502
FT CA_BIND 558 566
FT CA_BIND 620 628
FT DISULFID 76 86
FT DISULFID 666 675
FT DISULFID 681 736
FT DISULFID 789 795
FT CARBOHYD 98 98
FT CARBOHYD 234 234
FT CARBOHYD 336 336
FT CARBOHYD 364 364
FT CARBOHYD 733 733
FT CARBOHYD 763 763
FT CARBOHYD 839 839
FT CARBOHYD 921 921
FT CARBOHYD 1011 1011
FT CARBOHYD 1018 1018
FT CARBOHYD 1039 1039
FT CONFLICT 844 844
FT CONFLICT 909 909
FT CONFLICT 926 926
SQ SEQUENCE 1167 AA; 127573 MW; A67D3A1C25C1AE40 CRC64;

```

```

Query Match 100.0%; Score 1025; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 7.6e-85;
Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GICARVDASFQPGSLAPTAQRCTYMDVIVLDGNSIYPMSEVOTFLRLVGLKFLDP 60
DB 140 GICARVDASFQPGSLAPTAQRCTYMDVIVLDGNSIYPMSEVOTFLRLVGLKFLDP 199
QY 61 EGIQVGLVYGGSPVHMSLGDPRTEEVYRAKNLSREGRRTKTAQAIMACTGEGSQ 120
DB 200 EGIQVGLVYGGSPVHMSLGDPRTEEVYRAKNLSREGRRTKTAQAIMACTGEGSQ 259
QY 121 SHGGRPEARLVLVVDGSHDEELPALMKACEAGRVYGLAVGHLRQRDPSSFL 180
DB 260 SHGGRPEARLVLVVDGSHDEELPALMKACEAGRVYGLAVGHLRQRDPSSFL 319
QY 181 REIRTIASDPDERFFENV 198
DB 320 REIRTIASDPDERFFENV 337

```

RESULT 2  
ITAH\_HUMAN

```

ID ITAH_HUMAN STANDARD; PRT; 1189 AA.
AC 09UKX5; O9UK01;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-11 precursor.
GN ITGA11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=fetal heart, and Osteoblast;
RX MEDLINE=99417678; PubMed=10486209;
RA Lehnert K., Ni J., Leung F., Gough S.M., Weaver A., Yao W.P., Liu D.,
RA Wang S.-X., Morris C.M., Krissansen G.W.;
RT "Cloning, sequence analysis, and chromosomal localization of the novel
RT human integrin alpha11 subunit (ITGA11).";
RT Genomics 60:179-187(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=fetal muscle, and uterus;
RX MEDLINE=99395147; PubMed=10464311;
RA Velling T., Kusche-Gullberg M., Seijersen T., Gullberg D.;
RA "CDNA Cloning and Chromosomal Localization of Human alpha(11)
RT Integrin. A collagen-binding, i domain-containing, beta(1)-associated
RT integrin alpha-chain present in muscle tissues.";
RT J. Biol. Chem. 274:25735-25742(1999).
RN [3]
RP SEQUENCE OF 954-1188 FROM N.A.
RC TISSUE=fibroblast;
RA Andreu N., Estivill X., Escarceller M., Sunmy L.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11
CC ASSOCIATES WITH BETA-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND
CC HEART. INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO
CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO
CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS,
CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN
CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.
CC -1- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING
CC FETAL MUSCLE CELLS (IN VITRO).
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWF_A DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWF_A DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP-REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; AF109681; AAF01258.1; -
DR EMBL; AF137378; AAD51919.2; -
DR EMBL; AL359064; CAB94392.1; -
DR HSSP; P17301; IAOX.
DR Gene; HGNC:6136; ITGA11.
DR MIM; 604789; -.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR Pfam; PF01839; FG-GAP 5.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.

```



```

Db      257 -GGKRGKAKKMMIVITDGEHSDSPDLKVKIQQSRDNVTRAVAVLGVYNRGINPETFL 314
Qy      181 REFTTASDPDERFFENV 198
      ||:|||||:|
Db      315 NEIKYIASDPDKKFFNV 332

RESULT 3
ITAL CHICK STANDARD; PRT; 285 AA.
AC Q09015;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-1 (laminin and collagen receptor) (VLA-1) (CD49a)
DE (Fragment).
DE ITGAL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gizzard;
RX MEDLINE=94357930; Pubmed=7521332.
RA Kern A., Britsewicz R., Bank I., Marcantonio E.E.;
RT "The role of the I domain in ligand binding of the human integrin
  alpha 1 beta 1.";
RL J. Biol. Chem. 269:22811-22816(1994).
CC -1- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC E-R IN COLLAGEN.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC ASSOCIATES WITH BETA-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10114; AAA59067.1; -.
DR HSSP; P17301; IAOX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR020355; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00327; VWF; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA_PARTIAL.
DR KWM Integrin; PS50234; VWFA; 1.
DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane.
FT DOMAIN 1 1
FT DOMAIN <1 >285 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 66 279 VWFA.
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 285 AA; 31503 MW; 1B05D3246CD5CA7E CRC64;

Query Match 46.1%; Score 473; DB 1; Length 285;
Best Local Similarity 45.5%; Pred. No. 1,2e-35;
Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;

1 GICARVDASRPOGSLPTARQRCCTTMDVYIVLDGNSIYPMWSEVQTFRLRVLQKLFDP 60

```

```

Db      36 GVCNSVSTFETKAVAPSVQECKTOLDIVIVLDGNSISYPMBSVAFINSLIRNMDIGP 95
Qy      61 EOIQVGLVQYESPVHEWSLGDFTREEVVRANKLSRREGRTKTAQAIWVACTGFSQ 120
Db      96 QOTQVIGVQGTVVEHFYNTYSTEEVMDALRLRQGTGTMTALGIDTAREEAFTE 155
Qy      121 SHGCRPEARLLVVTVDGESHDEELPAALKACEAGRVTRYGIATVGLHYLRORDDPSFL 180
Db      156 AHGARRGVQKVMIVITVDGESHNDYRQOEVIDKCEDENIQRFALILGYSRGNLSTEFV 215
Qy      181 REIRTIASDPDERFFENV 198
Db      216 ERIKSIASEPTEKHFENV 233

RESULT 4
ID      ITAI_HUMAN      STANDARD;      PRT; 1151 AA.
AC      P56199;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a)
GN      ITGA1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      SEQUENCE FROM N.A.
RX      MEDLINE=93155124; PubMed=8428973;
RA      Briesewitz R., Epstein M.R., Marcantonio E.E.;
RT      "Expression of native and truncated forms of the human integrin alpha
RT      1 subunit.";
RL      J. Biol. Chem. 268:2989-2996(1993).
CC      -1- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC      COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC      E-R IN COLLAGEN.
CC      -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC      ASSOCIATES WITH BETA-1.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC      WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC      -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC      -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC      -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC      -1- DATABASE: NAME=PROV; NOTE=CD guide CD49a entry;
CC      WWW="http://www.ncbi.nlm.nih.gov/ProV/cd/cd49a.htm".
DR      HSSP; P17301; IAOX.
DR      Gene; HGNC:6134; ITGA1.
DR      MIM: 192968; -.
DR      InterPro; IPR000413; Integrin_alpha.
DR      InterPro; IPR002035; VWF_A.
DR      Pfam; PFO0092; vwa; 1.
DR      Pfam; PFO0357; Integrin_A; 1.
DR      Pfam; PFO1839; FG-GAP; 5.
DR      PRINTS; PR00453; VWFADOMAIN.
DR      SMART; SM00191; Int_alpha; 5.
DR      SMART; SM00327; VWFA; 1.
DR      PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR      PROSITE; PS50234; VWFA; 1.
KW      Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW      Repeat; Calcium; Magnesium.
FT      DOMAIN 1 1113 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 1114 1136 POTENTIAL.
FT      DOMAIN 1137 1151 CYTOPLASMIC (POTENTIAL).
FT      REPEAT 16 75 FG-GAP 1.
FT      REPEAT 7 75 FG-GAP 2.
FT      DOMAIN 147 360 VWFA.
FT      REPEAT 349 404 FG-GAP 3.
FT      REPEAT 405 457 FG-GAP 4.
FT      REPEAT 459 520 FG-GAP 5.

```

```

FT      REPEAT 540 599 FG-GAP 6.
FT      REPEAT 602 654 FG-GAP 7.
FT      CA_BIND 470 478 POTENTIAL.
FT      CA_BIND 552 560 POTENTIAL.
FT      CA_BIND 614 622 POTENTIAL.
FT      SITE 1139 1142 GEFKR MOTIF.
FT      DISULFID 54 64 BY SIMILARITY.
FT      DISULFID 660 669 BY SIMILARITY.
FT      DISULFID 675 728 BY SIMILARITY.
FT      DISULFID 780 786 BY SIMILARITY.
FT      DISULFID 850 858 BY SIMILARITY.
FT      DISULFID 1002 1034 BY SIMILARITY.
FT      DISULFID 1037 1044 BY SIMILARITY.
FT      CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 812 812 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 855 855 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1074 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1085 1085 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 1151 AA; 127837 MW; 6B3F3C1AABF52808 CRC64;

Query Match      45.6%; Score 467.5; DB 1; Length 1151;
Best Local Similarity 46.5%; Pred. No. 2,1e-34;
Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

Qy      1 GICARVDSFQPGSLAPTAQRCPYMDVIVVLGNSIYPMBSVQFTRRLVGLFDP 60
Db      118 GICSVSPFPQVNVSIAP-VEECSTQDLDIVIVLDGNSISYPMBSVTAFLNDLKRMDIGP 176
Qy      61 EOIQVGLVQYESPVHEWSLGDFTREEVVRANKLSRREGRTKTAQAIWVACTGFSQ 120
Db      177 KOTQVIGVQGENYTHEHNLNKSSTEBVLVAKKIVRGGRGTMTALGDTARKEAFTE 236
Qy      121 SHGCRPEARLLVVTVDGESHDEELPAALKACEAGRVTRYGIATVGLHYLRORDDPSFL 180
Db      237 ARGARRGVQKVMIVITVDGESHNDYRQOEVIDKCEDENIQRFALILGYSRGNLSTEFV 296
Qy      181 REIRTIASDPDERFFENV 198
Db      297 ERIKSIASEPTEKHFENV 314

RESULT 5
ID      ITAI_RAT      STANDARD;      PRT; 1180 AA.
AC      P18614;
DT      01-NOV-1990 (Rel. 16, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Integrin alpha-1 precursor (Laminin and collagen receptor) (VLA-1)
DE      (CD49a).
GN      ITGA1.
OS      Rattus norvegicus (Rat).

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCBI\_TaxId=10116;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=90338125; PubMed=2380249;  
 CC Ignatius M.J., Large T.H., Houde M., Tawil J.W., Barton A.,  
 CC Esch F., Carbonetto S., Reichardt L.F.;  
 CC "Molecular cloning of the rat integrin alpha 1-subunit: a receptor  
 CC for laminin and collagen."  
 CC J. Cell Biol. 111:709-720(1990).  
 CC [2]  
 CC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.  
 CC MEDLINE=99313197; PubMed=10386626;  
 CC Nole M., Pepinsky R.B., Vanyamou S.Y., Kotellansky V.,  
 CC Gotwals P.J., Kapusas M.;  
 CC "Crystal structure of the alphabeta integrin I-domain: insights into  
 CC integrin I-domain function."  
 CC FEBS Lett. 452:379-385(1999).  
 CC -1- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND  
 CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-  
 CC E-R IN COLLAGEN.  
 CC -1- SUBUNIT: HETERO-DIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL, X52140; CAA36384.1; -.  
 CC DR PIR, A35854; A35854.  
 CC DR HSSP, P17301; IAOX.  
 CC DR InterPro, IPR000413; Integrin\_alpha.  
 CC DR InterPro, IPR002035; VWF\_A.  
 CC DR Pfam, PF000397; vwa; 1.  
 CC DR Pfam, PF000397; Integrin\_A; 1.  
 CC DR Pfam, PF01839; FG-GAP\_5.  
 CC DR PRINTS, PR00453; VWFADOMAIN.  
 CC DR SMART, SM00191; Int\_alpha; 5.  
 CC DR SMART, SM00327; VWA; 1.  
 CC DR PROSITE, PS00242; INTEGRIN\_ALPHA; 1.  
 CC DR PROSITE, PS50234; VWF\_A; 1.  
 CC DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 CC Signal; Repeat; Calcium; Magnesium.  
 CC KM SIGNAL. 1 28  
 CC FT CHAIN 29 1180 INTEGRIN ALPHA-1.  
 CC FT DOMAIN 29 1142 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 1143 1165 POTENTIAL.  
 CC FT DOMAIN 1166 1180 CYTOPLASMIC (POTENTIAL).  
 CC FT REPEAT 44 103 FG-GAP 1.  
 CC FT REPEAT 175 388 FG-GAP 2.  
 CC FT DOMAIN 175 388 VWF\_A.  
 CC FT REPEAT 377 432 FG-GAP 3.  
 CC FT REPEAT 433 484 FG-GAP 4.  
 CC FT REPEAT 485 565 FG-GAP 5.  
 CC FT REPEAT 567 626 FG-GAP 6.  
 CC FT REPEAT 629 681 FG-GAP 7.  
 CC FT CA\_BIND 497 505 POTENTIAL.  
 CC FT CA\_BIND 579 649 POTENTIAL.  
 CC FT CA\_BIND 641 649 POTENTIAL.  
 CC FT SITE 1166 1172 GPFKR MOTIF.  
 CC FT DISULFID 82 92 BY SIMILARITY.  
 CC FT DISULFID 687 696 BY SIMILARITY.

FT DISULFID 702 755 BY SIMILARITY.  
 FT DISULFID 807 813 BY SIMILARITY.  
 FT DISULFID 877 885 BY SIMILARITY.  
 FT DISULFID 1029 1062 BY SIMILARITY.  
 FT DISULFID 1066 1073 BY SIMILARITY.  
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 698 698 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 779 779 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 820 820 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 839 839 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 907 907 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 973 973 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1007 1007 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1103 1103 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1180 AA; 130808 MW; 8E5DA2BE02362BE4 CRC64;  
 Query Match 45.4%; Score 465.5; DB 1; Length 1180;  
 Best Local Similarity 46.5%; Pred. No. 3.3e-34;  
 Matches 92; Conservative 35; Mismatches 70; Indels 1; Gaps 1;  
 QY 1 GICARVDASFPQGSLATACRCPTMDVIVLDGNSITYWSEVQTLRLVGLKFLIDP 60  
 DB 146 GICSDVSPFPQVNVMSFAP-VOECSTQDIVIVLDGNSIYWSEVIAFLNLDLKMMDIGP 204  
 QY 61 BOIQVGLVQGESPPVHEMSLGDFTKEEVVAANKLSREGERETKTAQAIWVACGEFSO 120  
 DB 205 KOTQVIGVQGESNTHLENLKYSSTBEVLVAANKIGQGLQMTALGIDTPARKEATTE 264  
 QY 121 SHGGRPEARLVVTVGDSEHDCGLPALKACAGRTVRCIVLGHVLRORDPSSFL 180  
 DB 265 ARGARGVKXKVMVITVDESNDRLKQVIODCEDENIGRSIALIGHYNGNSTEFV 324  
 QY 181 REIRTIASDPDERFFNV 198  
 DB 325 EIKSIASEPTEKHFNV 342  
 RESULT 6  
 ITA2 MOUSE STANDARD; PRT; 1178 AA.  
 AC Q62469; Q62163;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)  
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6 x CBA; TISSUE=Lung;  
 RX MEDLINE=94363406; PubMed=8081889;  
 RA Edelman J.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,  
 RA Damjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;  
 RT "The mouse VLA-2 homologue supports collagen and laminin adhesion but

RT not virus binding";  
 RL Cell Adhes. Commun. 2:131-143 (1994).  
 RN [2]  
 RP SEQUENCE OF 450-1178 FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=94355691; PubMed=7521231;  
 RA Wu J.E., Santoro S.A.;  
 RT "Complex patterns of expression suggest extensive roles for the alpha  
 2 beta 1 integrin in murine development.";  
 RL Dev. Dyn. 199;292-314(1994).  
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A COLLAGEN RECEPTOR, BEING  
 CC RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER CELLS TO  
 CC COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE EXPRESSION,  
 CC FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
 CC EXTRACELLULAR MATRIX. IT IS ALSO A RECEPTOR FOR LAMININ, COLLAGEN  
 CC C-PROPEPTIDES AND E-CADHERIN. MICE HOMOCYGOUS FOR A NULL MUTATION  
 CC IN THE ALPHA-2 DIE VERY EARLY IN EMBRYOGENESIS.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, Z29987; CA82877.1; -;  
 DR EMBL, X75427; CA85178.1; -;  
 DR HSSP, P17301; IAOX.  
 DR MGD, MGI:96600; Itga2.  
 DR InterPro, IPR000413; Integrin\_alpha.  
 DR InterPro, IPR002035; VWF\_A.  
 DR Pfam, PF00092; vwa; 1.  
 DR Pfam, PF00357; Integrin\_A; 1.  
 DR Pfam, PF01839; FG-GAP; 5.  
 DR PRINTS, PR00453; VWFADOMAIN.  
 DR SMART, SMO0191; Int alpha; 5.  
 DR SMART, SMO0327; VWA\_1.  
 DR PROSITE, PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE, PS0234; VWFA; 1.  
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Platelet; Signal; Repeat; Calcium; Magnesium.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1178  
 FT DOMAIN 27 1129  
 FT TRANSMEM 1130 1151  
 FT DOMAIN 1152 1178  
 FT REPEAT 42 100  
 FT REPEAT ? ?  
 FT DOMAIN 185 375  
 FT REPEAT ? ?  
 FT REPEAT 431 483  
 FT REPEAT 485 546  
 FT REPEAT 548 607  
 FT REPEAT 612 664  
 FT CA\_BIND 496 568  
 FT CA\_BIND 560 568  
 FT CA\_BIND 624 632  
 FT SITE 480 482  
 FT SITE 1154 1158  
 FT DISULFID 80 89  
 FT DISULFID 677 734  
 FT DISULFID 786 792  
 FT DISULFID 862 873  
 FT DISULFID 1016 1047

FT DISULFID 1052 1057 BY SIMILARITY.  
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1054 1054 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B3C0240F465 CRC64;  
 Query Match 44.3%; Score 454; DB 1; Length 1178;  
 Best Local Similarity 46.0%; Pred. No. 3.6e-33;  
 Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;  
 QY 1 GICARVDASFOGSLAFTACRCPYMDVYVLDGNSIYVWSEVQTLRLVGLFTDP 60  
 DB 144 GICSDVSPDFQFLTFSFAVQACPSLDVWVVCDESNSIYWEAVKRLVVFVGLDIGP 203  
 QY 61 EGIQVGLVOGESPVHEWSLGDFTKXEVRAAKNLSRREGRETKQAQIWAQTEGFSQ 120  
 DB 204 KKTQVALIQYANEPRIINLNDPFTKEDVQATSETRHGGLDITVFAIFARDYAYSQ 263  
 QY 121 SHGGRPEARLLVVTDESHDGEELPAALACENAGRYGIAVLGHYLRORDPSSFL 180  
 DB 264 TSGGRPGATKVVVVTDESHDGSGLKTVIQCNDDIELRGIAVLGYLNNAIDTKLI 323  
 QY 181 REIRTIASDPDERFFNV 198  
 DB 324 KEIKAIASPTERYEFNV 341  
 RESULT 7  
 ITR2\_BOVIN  
 ID ITR2\_BOVIN STANDARD; PRT; 1170 AA.  
 AC P53710;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)  
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).  
 GN ITGA2.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94193647; PubMed=7511592;  
 RA Kamata T., Puzon W., Takada Y.;  
 RT "Identification of putative ligand binding sites within I domain of  
 RL Integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";  
 RJ J. Biol. Chem. 269:9659-9663 (1994).  
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,  
 CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT  
 CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN  
 CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER  
 CC CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE  
 CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
 CC EXTRACELLULAR MATRIX.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -





CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; X70793; CAA50064.1; -  
DR EMBL; X70792; CAA50063.1; -  
DR EMBL; X66198; CAA46928.1; -  
DR PIR; S22916; S22916.  
DR PDB; 1B9P; 25-FEB-99.  
DR PDB; 1B9Q; 25-FEB-99.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR003962; FNIII\_repeat.  
DR InterPro; IPR003129; TSPN.  
DR InterPro; IPR002035; WVF\_A.  
DR Pfam; PF00041; fn3; 8.  
DR Pfam; PF00092; wva; 2.  
DR Pfam; PF01391; Collagen; 4.  
DR Pfam; PF02210; TSPN; 1.  
DR PRINTS; PR00014; ENTPEP11.  
DR SMART; SM00453; WVFADOMAIN.  
DR SMART; SM00060; FN3; 7.  
DR SMART; SM00210; TSPN; 1.  
DR PROSITE; PS50234; WVFA; 2.  
KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Cell adhesion; Collagen; Glycoprotein; Signal; 3D-structure.  
FT SIGNAL 1 28  
FT CHAIN 1 1888  
FT DOMAIN 158 330 WVFA 1.  
FT DOMAIN 441 529 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 531 622 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 625 728 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 739 821 FIBRONECTIN TYPE-III 4.  
FT DOMAIN 829 922 FIBRONECTIN TYPE-III 5.  
FT DOMAIN 952 1010 FIBRONECTIN TYPE-III 6.  
FT DOMAIN 1042 1215 WVFA 2.  
FT DOMAIN 1227 1468 NONHELICAL REGION (NC4).  
FT DOMAIN 1469 1620 TRIPLE-HELICAL REGION 1.  
FT DOMAIN 1664 1786 TRIPLE-HELICAL REGION 2.  
FT DOMAIN 721 733 POLY-THR.  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SITE 1489 1491 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 1617 1619 CELL ATTACHMENT SITE (POTENTIAL).  
SQ SEQUENCE 1888 AA; 202666 MW; 39915BB9F46DD873 CRC64;

Query Match 22.2% Score 227.5; DB 1; Length 1888;  
Best Local Similarity 31.7%; Pred. No. 2e-12; Indels 17; Gaps 6;  
Matches 60; Conservative 37; Mismatches 75;

QY 12 PGGSLAPTAQRCPTWDVIVIVDGSNSI-YFMSEVQFLRLVGLKFLIDPQIOGLVQ 69  
DB 144 PEGNFTFC-KTPALADIIVLDGWSIGRFNRLVRLLENLVAFNNGSEKTVGLAQ 201  
QY 70 YGESVPEHWSLGDFTKEEVPAKKLSRREKRETKYCAIVACTEGFSQSHGRPEEA 129  
DB 202 YSGDRIEWHNLAVYKDAVLDVAVRLPYK-GGNLTIGALTLYILENSKRPAGARPGVS 260  
QY 130 RLIVVTVDGESGELPAALAKACEAGRTYGIIVLGHYLRQDPSSFLREIRTIASD 189  
DB 261 KIGIILITDKSD-DVIRPAKRLRDAG-TELPALGV-----KNADINELKELASE 308  
QY 190 PDREFFFNV 198  
DB 309 PDSTHYNV 317

RESULT 10  
CAIC\_CHICK

ID CAIC\_CHICK STANDARD; PRT; 3124 AA.  
AC P13944; Q04509;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 15-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Collagen alpha 1(XII) chain precursor (Fibrochimerin).  
GN COL12A1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=White Leghorn;  
RX MEDLINE=92011862; PubMed=1918137;  
RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,  
RA Niimura Y., Obara M., Kimata K.;  
RT "The complete primary structure of type XII collagen shows a chimeric  
RT molecule with reiterated fibronectin type III motifs, von Willebrand  
RT factor A motifs, a domain homologous to a noncollagenous region of  
RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp  
RT site."  
RL J. Cell Biol. 115:209-221 (1991).  
RN [2]  
RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND  
RP 2846-2873.  
RX MEDLINE=90062079; PubMed=2584192;  
RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;  
RT "Type XII collagen: A large multidomain molecule with partial  
RT homology to type IX collagen."  
RL J. Biol. Chem. 264:19772-19778 (1989).  
RN [3]  
RP SEQUENCE OF 2960-3076 FROM N.A.  
RX MEDLINE=87317590; PubMed=3476925;  
RA Gordon M.K., Gerecke D.R., Olsen B.R.;  
RT "Type XII collagen: distinct extracellular matrix component  
RT discovered by cDNA cloning."  
RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044 (1987).  
RN [4]  
RP SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.  
RC TISSUE=Embryo;  
RX MEDLINE=93042014; PubMed=1420368;  
RA Trueb J., Trueb B.;  
RT "The two splice variants of collagen XII share a common 5' end."  
RL Biochim. Biophys. Acta 1171:97-98 (1992).  
RN [5]  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=95370352; PubMed=7642694;  
RA Koch M., Bohrmann B., Mathison M., Hagios C., Trueb B., Chiquet M.;  
RT "Large and small splice variants of collagen XII: differential  
RT expression and ligand binding."  
RL J. Cell Biol. 130:1005-1014 (1995).  
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-  
CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE  
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE  
CC LOCALIZED IN THE PERITRABULAR MATRIX.  
CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF  
CC NONTRIPLE-HELICAL SEQUENCES.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE  
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER  
CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT  
CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYTIC. THE LARGE  
CC ISOFORM HAS MORE RESTRICTED EXPRESSION IN EMBRYONIC TISSUE THAN  
CC THE SMALL.  
CC -1- TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS,  
CC LIGAMENTS, PERICHOONDRIUM, AND PERIOSTEUM, ALL DENSE CONNECTIVE  
CC TISSUES CONTAINING TYPE I COLLAGEN.  
CC -1- DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS. TWO TRIPLE-  
CC HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPLE-HELICAL  
CC DOMAINS (NC1, NC2, AND NC3).  
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT



FT	DOMAIN	3086	3096	ASP/GLU-RICH (ACIDIC).
FT	BINDING	3111	3123	ARG/LYS-RICH (BASIC).
FT	BINDING	797	797	TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	890	890	TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	981	981	TO CHONDROITIN SULFATE (POTENTIAL).
FT	SITE	2438	2440	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	2899	2901	CELL ATTACHMENT SITE (POTENTIAL).
FT	CARBOHYD	32	32	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1006	1006	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1032	1032	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1044	1044	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1512	1512	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1767	1767	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2210	2210	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2273	2273	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2532	2532	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2683	2683	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	25	1188	T -> S (IN REF. 4).
FT	CONFLICT	1258	1258	D -> E (IN REF. 4).
FT	CONFLICT	1264	1264	P -> A (IN REF. 2).
FT	CONFLICT	2759	2759	L -> F (IN REF. 2).
FT	CONFLICT	2803	2803	V -> F (IN REF. 2).
FT	CONFLICT	2977	2977	OP -> AG (IN REF. 3).
FT	CONFLICT	3075	3076	
SE	SEQUENCE	3124 AA;	340578 MM;	0942858ADEF7F346CF CRC64;
Query Match		22.24;	Score 227.5;	DB 1;
Best Local Similarity		34.14;	Pred. No. 3,7e-12;	
Matches		59;	Conservative %32;	Indels 15; Gaps 5;
OY	28	DVAVLDGNSI--YPMSEVQFLRLAVGLFLTDPEDIOGVGVGSPSPHWSLGDFT	85	
DB	1199	DIVLVGVGSMVSGHPNPKTVANFLSRIVEVDIGDPDVQVGLQVSGDPRTENMLNAYRT	1258	
OY	86	KEEVVRAKUNLSRREGREYTAQAIWVACTGEGFSQSHGSGPEARLLVWVTDGESHGEE	145	
DB	1259	KEALLDVAITNLPYK-GGNTLTGMALEDILKKNFROEGRLPRARKIGVLLTIDGSGQDVV	1317	
OY	146	LPAALAKCEAGRVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNV	198	
DB	1318	TPSRRLADEG--VELVAIGI-----KNADEN-----ELKQIADPPDDIHAYNV	1358	
RESULT 11				
CAMA_CHICK	STANDARD;	PRT;	493 AA.	
AC	P05099;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cartilage matrix protein precursor (Matrilin-1).			
DB	MATN1 OR CMP.			
GN	Gallus gallus (Chicken).			
OS	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
NC	NCBI_TaxID=9031;			
OX	[1]			
RN	SEQUENCE FROM N.A.. AND SEQUENCE OF 24-43.			
RP	STRAIN=White Leghorn;			
RC	MEDLINE=89255246; PubMed=2542265;			
RX	Kiss I., Deak F., Holloway R.G. Jr., Delius H., Neubst K.A.,			
RA	Frimberger E., Atgraves W.S., Tesons P.A., Winterbottom N.,			
RA	Goelink P.F.;			
RT	"Structure of the gene for cartilage matrix protein, a modular			
RT	protein of the extracellular matrix. Exon/intron organization,			
RT	unusual splice sites, and relation to alpha chains of beta 2			
RT	integrins, von Willebrand factor, complement factors B and C2, and			
RT	epidermal growth factor."			
RL	J Biol. Chem. 264:8126-8134(1989).			
RN	[2]			
RP	SEQUENCE OF 78-493 FROM N.A.			
RX	MEDLINE=87092429; PubMed=3025875;			

RA Argaves W.S., Deak F., Sparks K.J., Kiss I., Goetinck P.F.;  
RT "Structural features of cartilage matrix protein deduced from cDNA";  
RN Proc. Natl. Acad. Sci. U.S.A. 84:464-468(1987).  
RP [3]  
RX MEDLINE=97406913; PubMed=9260286;  
RA Williams R., Kammerer R.A., Dames S.A., Schulthess T.,  
RT Blomere M.J., Engel J., Alexandrescu A.T.;  
RT "Heteronuclear NMR assignments and secondary structure of the coiled  
RT coil trimerization domain from cartilage matrix protein in oxidized  
RT and reduced forms";  
RT Protein Sci. 6:1734-1745(1997).  
CC -1- FUNCTION: CARTILAGE MATRIX PROTEIN IS A MAJOR COMPONENT OF THE  
CC EXTRACELLULAR MATRIX OF NONARTICULAR CARTILAGE. IT BINDS TO  
CC COLLAGEN.  
CC -1- SUBUNIT: HOMOTRIMER.  
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 2 VWFA DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X12353; CAA30915.1; -  
DR EMBL; X12346; CAA30915.1; JOINED.  
DR EMBL; X12347; CAA30915.1; JOINED.  
DR EMBL; X12348; CAA30915.1; JOINED.  
DR EMBL; X12349; CAA30915.1; JOINED.  
DR EMBL; X12350; CAA30915.1; JOINED.  
DR EMBL; X12351; CAA30915.1; JOINED.  
DR EMBL; X12352; CAA30915.1; JOINED.  
DR EMBL; M14792; AAA48695.1; -  
DR EMBL; M97497; AAC18872.1; -  
DR PIR; A33809; A33809.  
DR PIR; A26364; A26364.  
DR PDB; 1A05; 11-FEB-98.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF00092; VWF; 2.  
DR Pfam; PF00092; VWF; 2.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00181; EGF; 1.  
DR SMART; SM00327; VWF; 2.  
DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS50234; VWF; 2.  
KW EGF-like domain; Signal; Glycoprotein; Cartilage; Repeat;  
KW Coiled coil; 3D-structure.  
FT SIGNAL 1 23  
FT CHAIN 24 493 CARTILAGE MATRIX PROTEIN.  
FT DOMAIN 24 220 VWFA 1.  
FT DOMAIN 221 261 EGF-LIKE.  
FT DOMAIN 262 450 VWFA 2.  
FT DOMAIN 462 492 COILED COIL.  
FT CAROHD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DISULFID 33 219 POTENTIAL.  
FT DISULFID 225 236 BY SIMILARITY.  
FT DISULFID 232 245 BY SIMILARITY.  
FT DISULFID 247 260 BY SIMILARITY.  
FT DISULFID 263 449 POTENTIAL.  
SQ SEQUENCE 493 AA; 54034 MW; E4D3DACFA2B96A4 CRC64;  
Query Match 20.9%; Score 214; DB 1; Length 493;  
Best Local Similarity 33.3%; Pred. No. 6; Se-12;  
Matches 58; Conservative 35; Mismatches 63; Indels 18; Gaps 6;

QY 85 TKEVRAAKNLRRREGRTQAQIMVACTGSGHGRPEARLLVVTGESHGE 144  
DB 331 NKDIDKAIVKKAAYME-KGTWTGQALKXLYVSSFSFIANGARPGVKGIVFDGSSQD-Y 388  
QY 145 ELPAALKACENG-RVTRYGIAVLGHYLRQRDPSSFLREITIASDP-DEPF 196  
DB 389 ITDAKKAKADIGFRFVAVGVG-----NAVEDELREIASEPVAHEHYF 430  
RESULT 12  
CALC\_MOUSE STANDARD; PRT; 3119 AA.  
ID AC 060847; P70322;  
DT 15-JUN-1998 (Rel. 36; Created)  
DT 15-JUN-2002 (Rel. 41; Last sequence update)  
DT 15-JUN-2002 (Rel. 41; Last annotation update)  
DE Collagen alpha 1(XII) chain precursor.  
GN COL12A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-1 AND  
RP XI1B-1).  
RC STRAIN=Swiss Webster, and C57BL/6J; TISSUE=Skin;  
RX MEDLINE=96170761; PubMed=8601036;  
RA Boehme K., Li Y., Oh P.S., Olsen B.R.;  
RT "Primary structure of the long and short splice variants of mouse  
RT collagen XII and their tissue-specific expression during embryonic  
RT development";  
RL dev. Dyn. 204:432-445(1995).  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-2  
RP AND XI1B-2).  
RC STRAIN=C57BL/6J; TISSUE=Skin fibroblast;  
RX MEDLINE=99348349; PubMed=10419532;  
RA Kanita A.M., Reichenberger E., Baur S.T., Karimbux N.Y., Taylor R.W.,  
RA Olsen B.R., Nishimura I.;  
RT "Structural variation of type XII collagen at its carboxyl-terminal  
RT NCI domain generated by tissue-specific alternative splicing";  
RL J. Biol. Chem. 274:22053-22059(1999).  
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-  
CC CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE  
CC SURFACE OF THE FIBRILS. AND THE COL2 AND NC3 DOMAINS MAY BE  
CC LOCALIZED IN THE PERIFIBRILAR MATRIX (BY SIMILARITY).  
CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF  
CC NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).  
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; XI1A-1 (SHOWN HERE), XI1A-  
CC 2/ERR#, XI1B-1 AND XI1B-2; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC THE FINAL TISSUE FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OR  
CC ANY COMBINATION OF THE VARIOUS ISOFORMS.  
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TENDONS, PERICHONDRUM,  
CC SKIN, CORNEA, SCLERA, BLOOD VESSELS, AND PERIOSTEUM.  
CC -1- DEVELOPMENTAL STAGE: THE LONG NC3 XI1A ISOFORMS ARE PREDOMINANT AT  
CC EARLY STAGES (ED7 AND 11); AT LATER STAGES OF DEVELOPMENT (ED15  
CC AND 17) THE SHORT NC3 XI1B FORMS BECOME THE MAJOR FORMS. AS THE  
CC SHORT NC3 FORMS BECOME THE MAJOR PRODUCT, THE LONG SPLICED VARIANT  
CC CONTINUES TO BE EXPRESSED IN SEVERAL TISSUES, EVEN AFTER BIRTH.  
CC THE LONG NC1 ISOFORMS, XI1A-1 AND XI1B-1, PEAK IN 15-DAY OLD  
CC EMBRYOS AND DECREASE IN 17-DAY OLD ONES. THE EXPRESSION OF THE  
CC SHORT NC1 FORM XI1B-2 REMAINS CONSTANT THROUGHOUT LATE STAGES OF  
CC EMBRYONIC DEVELOPMENT (ED15 AND ED17).  
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT  
CC EACH END (BY SIMILARITY).  
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY  
CC SIMILARITY).  
CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY  
CC ISOFORM XI1A-2 IS A PROTEOGLYCAN.  
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
CC INTERRUPTED HELICES (FACIT) FAMILY.



OR THE SHORTER ISOFORM OR ANY COMINATION OF LONG AND SHORT ISOFORM CHAINS.

- TISSUE SPECIFICITY: FOUND IN COLLAGEN 1-CONTAINING TISSUES: BOTH SHORT AND LONG ISOFORMS APPEAR IN ANNIION, CHORION, SKELETAL MUSCLE, SMALL INTESTINE, AND IN CELL CULTURE OF DERMAL FIBROBLASTS, KERATINOCYTES, AND ENDOTHELIAL CELLS. ONLY THE SHORT ISOFORM IS FOUND IN LUNG, PLACENTA, KIDNEY, AND A SQUAMOUS CELL CARCINOMA CELL LINE.

- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT EACH END (BY SIMILARITY).

- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY SIMILARITY).

- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY THE LONG VARIANT IS A PROTEOGLYCAN.

- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.

- SIMILARITY: CONTAINS 4 WFPA DOMAINS.

- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: U73778; AAC51244.1; -

EMBL: U73779; AAD0483.1; -

HSSP: P02751; 1TTF.

Genew: HGNC:2188; COL12A1.

MIM: 120320; -

InterPro: IPR000087; Collagen.

InterPro: IPR003961; FN\_III.

InterPro: IPR003962; FNIII\_Repeat.

InterPro: IPR003129; TSPN.

InterPro: IPR002035; WF\_A.

Pfam: PF00041; fn3; 18.

Pfam: PF00092; wfa; 4.

Pfam: PF01391; Collagen; 4.

Pfam: PF02210; TSPN; 1.

PRINTS: PR00014; FNYPEIII.

PRINTS: PR00453; WFPADOMAIN.

SMART: SM00060; FN3; 17.

SMART: SM00210; TSPN; 1.

SMART: SM00327; WFA; 4.

PROSITE: PS50234; WFPA; 4.

Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.

POTENTIAL.

COLLAGEN ALPHA 1 (XII) CHAIN.

CHAIN 1 24

FT CHAIN 25 3063

FT DOMAIN 25 114

FT DOMAIN 140 316

FT DOMAIN 333 426

FT DOMAIN 440 616

FT DOMAIN 630 721

FT DOMAIN 722 812

FT DOMAIN 813 903

FT DOMAIN 904 998

FT DOMAIN 999 1085

FT DOMAIN 1086 1178

FT DOMAIN 1179 1371

FT DOMAIN 1384 1473

FT DOMAIN 1474 1564

FT DOMAIN 1565 1652

FT DOMAIN 1654 1751

FT DOMAIN 1752 1842

FT DOMAIN 1843 1932

FT DOMAIN 1933 2023

FT DOMAIN 2024 2114

FT DOMAIN 2115 2202

FT DOMAIN 2203 2291

FT DOMAIN 2323 2496

FT DOMAIN 2451 2746

FT DOMAIN 2747 2898

FT DOMAIN 2899 2941

FT DOMAIN 2942 3044

FT DOMAIN 3045 3063

FT BINDING 798 798

FT BINDING 889 889

FT BINDING 981 981

FT SITE 862 864

FT SITE 2779 2781

FT SITE 2895 2897

FT MOD\_RES 2944 2944

FT MOD\_RES 2947 2947

FT MOD\_RES 2950 2950

FT MOD\_RES 2952 2952

FT MOD\_RES 2959 2959

FT MOD\_RES 2965 2965

FT MOD\_RES 2968 2968

FT MOD\_RES 2971 2971

FT MOD\_RES 2983 2983

FT MOD\_RES 3000 3000

FT MOD\_RES 3003 3003

FT MOD\_RES 3014 3014

FT MOD\_RES 3023 3023

FT MOD\_RES 3026 3026

FT MOD\_RES 3029 3029

FT CARBOHYD 760 700

FT CARBOHYD 1763 1763

FT CARBOHYD 2206 2206

FT CARBOHYD 2528 2528

FT CARBOHYD 2679 2679

FT VARSPLIC 25 1188

SEQ SEQUENCE 3063 AA; 333189 MW; 75FEA78FA8E48293 CRC64;

Query Match 20.4%; Score 209.5; DB 1; Length 3063;

Best Local Similarity 31.8%; Pred. No. 1.6e-10;

Matches 55; Conservative 33; Mismatches 70; Indels 15; Gaps 4;

QY 28 DVVIVLDSGNSI--YPMSEVQFLRLVGVKLFIDPEQIOVGLVGVSGSPVHMSIGDRT 85

DB 1199 DIVLVDSWSIGRANFRTVVSFISRIVEVPDIGKRVQVLAQSGDPRTWQINARD 1258

QY 86 KEEVVRANKLSRRGRSEKTAQAIMVACTGFSGSHGGRPEARLIVVVDGESHGDEE 145

DB 1259 KKSLLQAVANLPYK-GGNTLTGMALNFRQONFRQAGMRPRARIGVILIDGKSODVE 1317

QY 146 LPALKACEAGRVYRGIAVLGHVLRQRDPSSFLREIRITASDPERFFENV 198

DB 1318 APSKLLKDEGVLELPAIGI-----KNADEV---ELKMLATDDDDHNVV 1358

RESULT 14

ID CANA HUMAN STANDARD; PRT; 496 AA.

AC P21941;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cartilage matrix protein precursor (Matrilin-1).

GN MATN1 OR CRTM OR CMP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE=91060568; PubMed=2246248;

RA Jenkins R.N., Osborne-Lawrence S.L., Sinclair A.K., Eddy R.L. Jr.,

RA Byers M.G., Snows T.B., Duby A.D.,

RT "Structure and chromosomal location of the human gene encoding

cartilage matrix protein.";

```

RL J. Biol. Chem. 265:19624-19631(1990).
CC -1- FUNCTION: CARTILAGE MATRIX PROTEIN IS A MAJOR COMPONENT OF THE
CC EXTRACELLULAR MATRIX OF NONARTICULAR CARTILAGE. IT BINDS TO
CC COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 WFMA DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M55682; AAB38702.1; -.
DR EMBL, M55675; AAB38702.1; JOINED.
DR EMBL, M55676; AAB38702.1; JOINED.
DR EMBL, M55677; AAB38702.1; JOINED.
DR EMBL, M55679; AAB38702.1; JOINED.
DR EMBL, M55680; AAB38702.1; JOINED.
DR EMBL, M55681; AAB38702.1; JOINED.
DR EMBL, M55683; AAA63904.1; ALT_SEQ.
DR PIR, A37979; A37979.
DR HSSP, P05099; IAOS.
DR Genev, HGNC:6907; MATN1.
DR MIM, 115437; -.
DR InterPro, IPR000561; EGF-like.
DR InterPro, IPR002035; WVF_A.
DR Pfam, PF00008; EGF_1.
DR Pfam, PF00092; Wvf_2.
DR PRINTS, PRO0453; WVFADOMAIN.
DR SMART, SM00181; EGF_1.
DR SMART, SM00327; WVF_2.
DR PROSITE, PS00022; EGF_1; FALSE_NEG.
DR PROSITE, PS01186; EGF_2; 1.
DR PROSITE, PS50234; WVF_2.
KW EGF-like domain; signal; Glycoprotein; Cartilage; Repeat; Coiled coil.
FT SIGNAL 1 222
FT CHAIN 23 496
FT DOMAIN 23 222
FT DOMAIN 223 263
FT DOMAIN 264 453
FT DOMAIN 467 495
FT CARBOHYD 76 76
FT DISULFID 35 221
FT DISULFID 227 238
FT DISULFID 234 247
FT DISULFID 249 262
FT DISULFID 265 452
SQ SEQUENCE 496 AA; 53700 MW; 2D880A8114C7940F CRC64;

Query Match 20.0%; Score 205; DB 1; Length 496;
Best Local Similarity 33.5%; Pred. No. 4,3e-11;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

```

```

AC Q91145;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain (Fragment).
OS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
OC Notophthalmus.
OX NCBI_Taxid=8316;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95246925; Pubmed=7729585;
RA Wei Y., Yang E.V., Klatz K.P., Taassava R.A.;
RT "Monoclonal antibody Mr2 identifies the urodele alpha 1 chain of type
RT XII collagen, a developmentally regulated extracellular matrix
RT protein in regenerating newt limbs."
RL Dev. Biol. 168:503-513(1995).
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS, THE COIL DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COI2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY). COULD PLAY
CC A DEVELOPMENTAL ROLE IN REGENERATION.
CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
CC NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).
CC -1- DEVELOPMENTAL STAGE: EXPRESSION STARTS AT 3 DAYS AFTER AMPUTATION
CC IN CELLS OF THE BASAL LAYER OF THE WOUND EPITHELIUM. AT DAY 10,
CC EXPRESSION IS FOUND IN BOTH THE BASAL WOUND EPITHELIAL CELLS AND
CC THE DISTAL MESenchyme CELLS. AT MID-BUD AND LATE-BUD BLASTEMA
CC STAGES, WOUND EPITHELIUM EXPRESSION HAS DECREASED, WHEREAS THE
CC MESenchyme REMAINS STRONGLY ACTIVE IN TRANSCRIPTION AND SHOWED A
CC TENDENCY TOWARD DISTAL REGIONALIZATION. CONDENSING CARTILAGE SHOWS
CC NO SIGNAL. FINALLY, AT THE LATE DIGIT STAGE, EXPRESSION BECOMES
CC LARGELY RESTRICTED TO THE PERICHONDRIUM.
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END (BY SIMILARITY).
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -1- SIMILARITY: CONTAINS 2 WFMA DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U19494; AAB0217.1; -.
DR HSSP, P02751; IENA.
DR InterPro, IPR003961; FN_III.
DR InterPro, IPR002035; WVF_A.
DR Pfam, PF00041; fn3; 7.
DR Pfam, PF00092; Wvf_2.
DR SMART, SM00050; FN3; 5.
DR SMART, SM00327; WVF_1.
DR PROSITE, PS50234; WVF_2.
KW Extracellular matrix; Connective tissue; Repeat; Cell adhesion;
KW Collagen; Glycoprotein.
FT NON_TER 1 1
FT DOMAIN <1 49
FT DOMAIN 63 154
FT DOMAIN 155 245
FT DOMAIN 246 338
FT DOMAIN 339 432
FT DOMAIN 433 519
FT DOMAIN 520 612
FT DOMAIN 633 805
FT DOMAIN 818 907
FT DOMAIN 908 >929

```



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:02 ; Search time 20.3291 Seconds

(without alignments)  
2006.842 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025  
Sequence: 1 GICARVDASFQPGSLAFTA.....FLREIRTIASDPDERFFENV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	473	46.1	1171	13 Q42094	Q42094 gallus gall
2	336	32.8	191	6 Q29124	Q29124 sus scrofa
3	207.5	20.2	1472	13 Q90200	Q90200 gallus gall
4	205.5	20.0	1207	4 Q9B0U7	Q9B0U7 homo sapien
5	204.5	20.0	1140	4 Q9P218	Q9P218 homo sapien
6	199	19.4	3767	5 Q9UAI3	Q9UAI3 caenorhabdi
7	196	19.1	956	11 Q8R542	Q8R542 mus musculu
8	191	18.6	937	4 Q96FT5	Q96FT5 mus musculu
9	191	18.6	956	11 Q99K64	Q99K64 mus musculu
10	188.5	18.4	2104	5 Q21281	Q21281 caenorhabdi
11	188.5	18.4	2104	5 Q964N4	Q964N4 caenorhabdi
12	182.5	17.8	526	4 Q9Y3N2	Q9Y3N2 homo sapien
13	182.5	17.8	954	4 Q8MXV8	Q8MXV8 homo sapien
14	182.5	17.8	957	4 Q96P44	Q96P44 homo sapien
15	182	17.8	387	11 Q60863	Q60863 mus musculu
16	181.5	17.7	957	4 Q9H0V3	Q9H0V3 homo sapien

17	181	17.7	549	6 Q02808	Q02808 bos taurus
18	177.5	17.3	360	4 Q9UGC3	Q9UGC3 homo sapien
19	177	17.3	664	4 Q19013	Q19013 erinaceus e
20	175	17.1	724	5 Q04588	Q04588 eimeria max
21	172.5	16.8	386	11 Q91YC3	Q91YC3 otomys ango
22	170.5	16.6	421	11 Q9UK16	Q9UK16 mus musculu
23	170.5	16.6	843	4 Q05707	Q05707 homo sapien
24	169	16.5	537	4 Q96AA0	Q96AA0 homo sapien
25	169	16.5	1169	11 Q9QXH4	Q9QXH4 mus musculu
26	168.5	16.4	200	11 Q88494	Q88494 mus musculu
27	168	16.4	1703	11 Q92019	Q92019 mus musculu
28	167	16.3	517	4 Q43853	Q43853 mus musculu
29	165.5	16.1	382	11 Q91Y33	Q91Y33 homo sapien
30	165.5	16.1	385	11 Q91YF6	Q91YF6 jaculus jac
31	164.5	16.0	410	11 Q92114	Q92114 micromys mi
32	163.5	16.0	380	11 Q91YH0	Q91YH0 allactaga e
33	163	15.9	1253	11 Q97566	Q97566 calomyscus
34	162.5	15.9	411	6 Q97566	Q97566 canis faml
35	161.5	15.8	417	11 Q920Q9	Q920Q9 inia geoffr
36	161	15.7	755	4 Q00261	Q00261 dryomys nit
37	159.5	15.6	385	11 Q91YH1	Q91YH1 homo sapien
38	159.5	15.6	385	11 Q91YH1	Q91YH1 cricetus
39	159.5	15.6	385	11 Q91YH1	Q91YH1 lophuromys
40	159.5	15.6	385	11 Q91YH1	Q91YH1 phodopus ro
41	159	15.5	393	6 Q9GL87	Q9GL87 tatera kemp
42	158.5	15.5	406	6 Q77672	Q77672 mesopiodon
43	158.5	15.5	421	11 Q9EPN7	Q9EPN7 scalopus aq
44	158	15.4	412	6 Q9MZE6	Q9MZE6 heliophobiu
45	158	15.4	1160	11 Q9R200	Q9R200 ziphius cay

## ALIGNMENTS

RESULT 1	Q42094	PREDIMINARY;	PRT;	1171 AA.
ID	Q42094			
AC	Q42094			
DT	01-JAN-1998	(TRENBLREL. 05, Created)		
DT	01-JAN-1998	(TRENBLREL. 05, Last sequence update)		
DT	01-MAR-2002	(TRENBLREL. 20, Last annotation update)		
DE	ALPHA1 integrin.			
OS	Gallus gallus (Chicken).			
OC	Gallus gallus			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=GITZARD;			
RX	MEDLINE=97476270; PubMed=9334246;			
RA	Obata H., Hayashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,			
RA	Sobue K.;			
RT	"Smooth muscle cell phenotype-dependent transcriptional regulation of			
RT	the alpha integrin gene.";			
RL	J. Biol. Chem. 272:26643-26651 (1997).			
DR	EMBL; AB000470; BAA23160.1; -			
DR	EMBL; AB000471; BAA23161.1; -			
DR	HSSP; P17301; IA0X.			
DR	InterPro; IPR000413; Integrin_alpha.			
DR	InterPro; IPR002035; VWF A.			
DR	Pfam; PF01839; FG-GAP; 5			
DR	Pfam; PF00357; Integrin_A; 1.			
DR	Pfam; PF00092; vwa; 1.			
DR	PRINTS; PR00453; VWFADOMAIN.			
DR	SMART; SM00191; Int_alpha; 4.			
DR	SMART; SM00327; VWA; 1.			
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.			
DR	PROSITE; PS50234; VWA; 1.			
KW	Integrin.			
SO	SEQUENCE	1171 AA;	130228 MW;	B505A4D65F09736E CRC64;

Query Match 46.1%; Score 473; DB 13; Length 1171;

Best Local Similarity 45.5%; Pred: No. 5.3e-35;  
Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;

```

QY 1GICAVVDASFOQGGSLAIPAORCEPTMDVDIVYLDGNSIYPMSEVOPTFLRLVGLKFLIPD 60
Db 135 G[:::SSSSSSFFFEVVKVAPBSVOEKTOLDIVYLDGNSIYPMSEVAFPLNSLRNMDIG 194
QY 61 EOIQVGLVOYGESPVHEWSLGDFRTEKEVEVRAKONLSREGRBETKTAQAIMVACTGFSQ 120
Db 195 QOTQYIGIYOGGTVLHFEFLNTRYSTTEEEMDALRLRQGGGQOTMALGIDTAREAFIE 254
QY 121 SHGEPPEARLLIVVYTDGESHGDEELPAALKACEAGRVTTRYGIAVYGHYLRORDSSFL 180
Db 255 AHGARGGVOKWVIVITDGEHSHNRYRLOEVIDKCEDENIORFAIAIIGSYSGNLSSTEKEV 314
QY 181 REIRTIASDPDERFFPNV 198
Db 315 EEIKSIASKPEYKHEFPNV 332

```

## RESULT 2

ID	Q29124	PRELIMINARY;	PTT;	191 AA.
AC	Q29124;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)			
DE	VLA-2 (Fragment).			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	TISSUE=ENDOTHELIAL CELLS;			
RX	MEDLINE=95036279; PubMed=7949129;			
RA	Bahou W.F., Potter C.L., Mirza H.;			
RT	"The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific			
RT	recognition sequence for endothelial cell attachment and spreading;			
RL	molecular and functional characterization.";			
RL	Blood 84:3734-3741(1994).			
DR	EMBL; Z12137; CAAT8125.1; -.			
DR	HSSP; P17301; IAOX.			
DR	InterPro; IPR002035; VMF_A.			
DR	Pfam; PF000092; vma; 1.			
DR	PRINTS; PR00453; VMFADOMAIN.			
DR	SMART; SM00327; vma; 1.			
DR	PROSITE; PSS0234; VMFA; 1.			
FT	NON_TER 1			
FT	NON_TER 191			
SO	SEQUENCE 191 AA; 20896 MW; 2E228BA72EC699D8 CRC64;			

RESULT 3  
Q90ZAO  
ID Q90ZAO PRELIMINARY; PRT; 1472 AA.  
AC Q90ZAO;

DT	01-DEC-2001	(TREMBLrel. 19, Created)
DT	01-DEC-2001	(TREMBLrel. 19, last sequence update)
DT	01-JUN-2002	(TREMBLrel. 21, last annotation update)
DE	Collagen type XX alpha 1 precursor.	
OS	Gallus gallus (chicken).	
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
CC	Gallus.	
OX	NCBI_TaxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21303548; PubMed=11274142;	
RA	Koch M., Foley J.E., Hahn R., Zhou P., Burgess R.E., Gerecke D.R.,	
RT	Gordon M.K.	
RT	"alpha 1(XX) Collagen, a New Member of the Collagen Subfamily, Fibril-	
RT	associated Collagens with Interrupted Triple Helices.";	
RL	J. Biol. Chem. 276:23120-23126(2001).	
DR	EMBL; AF312825; AAK58847.1; -	
DR	InterPro; IPR000087; Collagen.	
DR	InterPro; IPR003961; FN III.	
DR	InterPro; IPR003129; TSPN.	
DR	InterPro; IPR002035; VWF_A.	
DR	Pfam; PF01391; collagen; 4.	
DR	Pfam; PF00041; fn3; 6.	
DR	Pfam; PF02210; TSPN; 1.	
DR	Pfam; PF00092; wva; 1.	
DR	ProDom; PD000007; Collagen; 1.	
DR	PROSITE; PS50234; WVFA; 1.	
KW	Signal.	
FT	1 28 POTENTIAL.	
SQ	SEQUENCE 1472 AA; 156903 MW; 5361611579C56FFD CRC64;	

```

RESULT 4
Q9B0U7
ID Q9B0U7 PRELIMINARY; PRT; 1207 AA.
AC Q9B0U7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE BA261N11.4.1 (K1AA1510, isoform 1).
GN BA261N11.4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP
RA Hall R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121827; CAC36117.1; -.
DR HSSP; P17301; INOX.
DR InterPro; IPR000087; Collegen.

```

DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00041; Collagen\_1.  
 DR Pfam; PF00041; fn3; 6.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00060; FN3; 6.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00327; vwa; 1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 SQ SEQUENCE 1207 AA; 128085 MW; 3681586C6C44F1C1 CRC64;

Query Match 20.0%; Score 205.5; DB 4; Length 1207;  
 Best Local Similarity 29.5%; Pred. No. 3.5e-10;  
 Matches 59; Conservative 37; Mismatches 71; Indels 33; Gaps 7;

Qy 11 QPGSLAFTAQ-----RC--PTYMDVIVLDGNSI--YPMSEVQTFLR 50  
 Db 151 EPQVAFPTSDPRPTGSEWRETGPQFRCLPVPADWFLVDGWSIGSHFQVQKDFLA 210  
 Qy 51 RLVGKFLFDPEQIQVGLVQYGSFVHWSLGDPRFKKEVVAANKLSRREGRTTAQAI 110  
 Db 211 SVIAPFEIGPKVQVGLTQYSGDAQTEWDLNSLSTKEQVLAVRRL-RYKGGNTFTGLAL 269  
 Qy 111 MVACTGFSQSHGSPPEARLLVVTGDESHDGEELPALKACAGRYTRYGIAYLGHYL 170  
 Db 270 THVLGQNIQPAAGLRFPEAKVILVTGKSD--DVHTAARVLKDLGNVAVGV----- 322  
 Qy 171 RRQDDPSFLREIRTIASDP 190  
 Db 323 -KNADEA---ELRLLASPP 337

## RESULT 5

Q9P218 PRELIMINARY; PRT; 1140 AA.  
 AC Q9P218;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE KIAA1510 protein (Fragment).  
 GN KIAA1510.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RX MEDLINE=20277482; PubMed=10819331;  
 RP SEQUENCE FROM N.A.  
 RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human  
 RT genes.XVII.The complete sequences of 100 new cDNA clones from brain  
 RT which code for large proteins in vitro."  
 RL DNA Res. 7:143-150(2000).  
 DR EMBL; AB040943; BAA96034.1; -  
 DR HSSP; P17301; IAOX.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01391; Collagen\_2.  
 DR Pfam; PF00041; fn3; 5.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00060; FN3; 5.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00327; vwa; 1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 1140 AA; 120478 MW; 9EABD7D24FD67DAD CRC64;

Query Match 20.0%; Score 204.5; DB 4; Length 1140;  
 Best Local Similarity 32.4%; Pred. No. 4e-10;  
 Matches 56; Conservative 34; Mismatches 66; Indels 17; Gaps 6;

Qy 22 RC--PTYMDVIVLDGNSI--YPMSEVQTFRLRLVGKFLFDPEQIQVGLVQYGSFVH 77  
 Db 21 RCLPVPADWFLVDGWSIGSHFQVQKDFLASVIAPFEIGPKVQVGLTQYSGDAQTE 80  
 Qy 78 WSLGDFPKKEVVAANKLSRREGRTTAQAIWVACTGFSQSHGSPPEARLLVVTVD 137  
 Db 81 WDLNSLSTKEQVLAVRRL-RYKGGNTFTGLALTHVLGQNIQPAAGLRFPEAKVILVTD 139  
 Qy 138 GESHGDELPAALKACAGRYTRYGIAYLGHYLRQDDPSFLREIRTIASDP 190  
 Db 140 GKSD--DVHTAARVLKDLGNVAVGV-----KNADEA---ELRLLASPP 180

## RESULT 6

Q9UAL3 PRELIMINARY; PRT; 3767 AA.  
 ID Q9UAL3;  
 AC Q9UAL3; G21340;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Transmembrane cell adhesion receptor MUA-3 precursor (K08E5.3 protein)  
 DE (Fragment).  
 GN MUA-3 OR K08E5.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditiidae;  
 OC Rhabditiidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.;  
 RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.;  
 RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Lu Z., Vogel B., Hedgecock E.;  
 RT "mua-3 mRNA Splicing Pattern Revealed."  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.;  
 RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.;  
 RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Berkis M., Smith A., Lloyd C.R.;  
 RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF119060; AAD29428.1; -  
 DR EMBL; Z30974; CA83326.2; -  
 DR EMBL; Z30423; CA83326.2; JOINED.  
 DR EMBL; Z30423; CAC42345.1; -  
 DR EMBL; Z30974; CAC42345.1; JOINED.  
 DR HSSP; P01130; IIDL.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR001762; Disintegrin.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR002172; LDL\_recept\_A.  
 DR InterPro; IPR000882; SEA\_domain.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00008; EGF; 33.  
 DR Pfam; PF00057; Id1\_recept\_a; 3.  
 DR Pfam; PF01390; SEA; 2.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PR00289; DISINTEGRIN.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR PRINTS; PR00453; VWFADOMAIN.

```

DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_Like; 45.
DR SMART; SM00192; LDLA; 4.
DR SMART; SM00200; SEA; 2.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_32.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00068; LDLA_2; 2.
DR PROSITE; PS00024; SEA; 4.
DR PROSITE; PS00234; WFA; 2.
DR Calcium-binding; EGF-like domain; Glycoprotein; Receptor; Repeat;
KW Signal.
FT SIGNAL. 1 24 POTENTIAL.
FT CHAIN 25 >3767 3. TRANSMEMBRANE CELL ADHESION RECEPTOR MUA-
FT NON_TER 3767 3767
SQ SEQUENCE 3767 AA; 417284 MW; 8DA3A5E5A50A8B8E CRC64;

Query Match
Best Local Similarity 19.4%; Score 199; DB 5; Length 3767;
Matches 59; Conservative 33; Mismatches 83; Indels 28; Gaps 8;

QY 6 VDASFO---PQSLAFTAQRCPY--MDVYIVLDGNSIYPM--SEVQTFRLRYGKLF 58
DB 1204 VDVSSNANLPGRVCTVQTTCPKQKTDLVFLIDSGSIGSYFKNEVLFAVEFELFEI 1263
QY 59 DEBOIOVGLVOYGESPVHEWSLGDFTKEEVVRAAKNLSRRREGRETKTAQIMVACTEGF 118
DB 1264 GRSKTRVGLIOYSDQIRHEFDLDQYGDRLSLKLGISETOYLTGL--TRTAAAIQHWQEGF 1322
QY 119 SQSHGRP---EAARLLVVVTGDESHDGEELPAALACAGR--VTRYGIVLGHYLR 172
DB 1323 SERGARFPQSDIARVAIITLDRSODNVTGPA-----DSARKLSINTPAIGVTDHVL-- 1375
QY 173 QRPDSFLREIRTIASDPDERFF 195
DB 1376 -----ASELESIASGPRWFF 1391

RESULT 7
ID Q8R542 PRELIMINARY; PRT; 956 AA.
AC Q8R542;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE Matrilin-2.
GN MATN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21841526; PubMed=11852232;
RA Matas L., Korpos E., Deak F., Liu Z., Beier D.R., Aszodi A., Kiss I.;
RT "Comparative analysis of the mouse and human genes (Matn2 and MATN2)
RT for matrilin-2, a filament-forming protein widely distributed in
RT extracellular matrices."
RT Matrix Biol. 21:163-174(2002).
DR EMBL; AF358844; AAM11539.1; -.
DR EMBL; AF358831; AAM11539.1; JOINED.
DR EMBL; AF358832; AAM11539.1; JOINED.
DR EMBL; AF358833; AAM11539.1; JOINED.
DR EMBL; AF358834; AAM11539.1; JOINED.
DR EMBL; AF358835; AAM11539.1; JOINED.
DR EMBL; AF358836; AAM11539.1; JOINED.
DR EMBL; AF358837; AAM11539.1; JOINED.
DR EMBL; AF358838; AAM11539.1; JOINED.
DR EMBL; AF358839; AAM11539.1; JOINED.

```

```

DR EMBL; AF358840; AAM11539.1; JOINED.
DR EMBL; AF358841; AAM11539.1; JOINED.
DR EMBL; AF358842; AAM11539.1; JOINED.
DR EMBL; AF358843; AAM11539.1; JOINED.
SQ SEQUENCE 956 AA; 106748 MW; 3E4A608FF92BDE55 CRC64;

Query Match
Best Local Similarity 19.1%; Score 196; DB 11; Length 956;
Matches 64; Conservative 33; Mismatches 83; Indels 32; Gaps 8;

QY 6 VDASFOQSGSLAFTAQRCPY-----MDVYIVLDGNSI--YPMSEVQTF 49
DB 21 VDGERPQAPRRPSSRRHRYMPQTALLSSCENKRAIDLVIIDSSRSVNTDYAKVKEFI 80
QY 50 RLTVKLEIDPEQIOVGLVOYGESPVHEWSLGDFTKEEVVRAAKNLSRRREGRETKTAQ 109
DB 81 LDLIQFLDIGPDVTRVGLQGSTVKEFSLKTRFRKSEVERAVAKRM-RHLSTGTMTGLA 139
QY 110 IMVACTEGFSQSHGRP---EAARLLVVVTGDESHDGEELPAALACAGRVTTRYGIVL 166
DB 140 IOYALNINFASEAGRPARENVPRILIMVTDGRQDSVAEVA-KARNTG-ILTFALGV- 196
QY 167 GHYLRQRPDSFLREIRTIASDPDERFFNV 198
DB 197 -----GQVD---INTLKAIGSEPHKDHVFLV 219

RESULT 8
ID Q96FT5 PRELIMINARY; PRT; 937 AA.
AC Q96FT5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Matrilin 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Strausberg R.;
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010444; AAH0444.1; -.
DR InterPro; IPR000152; ASX_hydroxyl.
DR InterPro; IPR000561; EGF_Like.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR002035; WVF_A.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00092; WVF; 2.
DR PROSITE; PS00103; ASX_HYDROXYL; UNKNOWN_9.
DR PROSITE; PS01186; EGF_2; UNKNOWN_9.
DR PROSITE; PS0234; WVF; 2.
SQ SEQUENCE 937 AA; 104775 MW; 7C76C5CE70F06E40 CRC64;

Query Match
Best Local Similarity 18.6%; Score 191; DB 4; Length 937;
Matches 58; Conservative 33; Mismatches 67; Indels 18; Gaps 7;

QY 28 DVVIVLDGNSI--YPMSEVQTFRLRYGKLFIDPEQIOVGLVOYGESPVHEWSLGDFT 85
DB 57 DLVFLIDSSRSVNHDAKVEFIVDLQFLDIGPDVTRVGLQYGSTVKEFSLKTRFR 116
QY 86 KEEVVRAAKNLSRRREGRETKTAQIMVACTEGFSQSHGRP---EAARLLVVVTGDESHD 142
DB 117 KSEVERAVAKRM-RHLSTGTMTGLAIOYALNINFASEAGARPLRENVPRIVIMVTDGRPOD 175
QY 143 GEEVPAALACAGRVTTRYGIVLGHYLRQRPDSFLREIRTIASDPDERFFNV 198
DB 176 SVAEVA-KARNTG-ILTFALGV-----GQVD---FNTLKSIGSEPHDHVFLV 219

```

```

RESULT 9
Q99K64 ID Q99K64 PRELIMINARY; PRT; 956 AA.
AC Q99K64;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Matrilin 2.
GN MATN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005429; AA05429.1; -.
DR HSSP; P05099; IA05.
DR MGD; MGI:109613; Matn2.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00092; vwa; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00181; EGF; 10.
DR SMART; SM00179; EGF_CA; 10.
DR SMART; SM00327; vwa; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 7.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS50234; VWFA; 2.
KW EGF-like domain; Glycoprotein; Hydroxylation.
SQ SEQUENCE 956 AA; 106706 MW; 373BFC74370BFEE CRC64;

Query Match 18.6%; Score 191; DB 11; Length 956;
Best Local Similarity 29.7%; Pred. No. 5,7e-09;
Matches 63; Conservative 34; Mismatches 83; Indels 32; Gaps 8;

QY 6 VDASFGPGSLAPTRQRCPT;-----MDVYIVLDGNSI--YPMSEVQFL 49
DB 21 VDGRRPQAPRPPSRGRHVRMPQOTALLSSCENKRAIDIVFIIDSSRSVNTHDYAKVKEFI 80
QY 50 RLVLKLTIDPEQIOVGLVOYGESPVHEMSLDGFRTEKEVRAAKNLSRREGRETKTQDA 109
DB 81 LDIDLPDIDGPDVTVKGLIQIGSTYKNEPSLTKTFKSKSEVERAVKRM-RHLSGTMTGLA 139
QY 110 IMVACTEGFSQSHGRP---EAARLLVVTDSHDEGELPALAKACEAGRTVRYGIAVL 166
DB 140 IQVALNINAFSEAEGRPLRENVPRITMITVDGRPDQSVAEVA--KARNTG-ILIPALGV- 196
QY 167 GHYLRQRPDSSFLREIRTIASDPDERPFENV 198
DB 197 -----GQVD---LNTLKAIGSEPHKHVFLV 219

RESULT 10
Q21281 ID Q21281 PRELIMINARY; PRT; 2104 AA.
AC Q21281;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 230.1 kDa protein.
GN K07D8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;

```

```

RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Favello T.;
RT "The sequence of C. elegans cosmid K07D8."
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; L16679; AA28092.5; -.
DR HSSP; P02468; ITLE.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR000082; SEA_domain.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00008; EGF; 20.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00181; EGF; 35.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00200; SEA; 4.
DR SMART; SM00327; vwa; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_15.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS50024; SEA; 4.
DR PROSITE; PS50234; VWFA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hypothetical protein;
KW Repeat.
SQ SEQUENCE 2104 AA; 230051 MW; C98CD462CC43F8B CRC64;

Query Match 18.4%; Score 188.5; DB 5; Length 2104;
Best Local Similarity 28.5%; Pred. No. 2.8e-08;
Matches 57; Conservative 33; Mismatches 85; Indels 25; Gaps 8;

QY 6 VDASFGPGSLAPTRQRCPT;MDVYIVLDGNSIYPM--SEVQFLRLVLKGLFI 58
DB 411 VDSSNNMLPGRGCTCTSTACPAQPTDLPFLIDSGSIGSVFQTEVLRFLAEFTLELDI 470
QY 59 DPEQIOVGLVOYGESPVHEMSLDGFRTEKEVRAAKNLSRREGRETKTQDAIMVACTEGF 118
DB 471 APQTRSVSVYSQIIRHEFGLDYSRKSQNAIRINIEYTG-TRTGAALIEVANAELAF 529
QY 119 SQSHGRP---EAARLLVVTDSHDEGELPALAKACEAGR--VTRYGIAVLGHYLRQR 174
DB 530 SERRGARPVGVSVAVLITDGRSQDNVTRSD-----NARQDILQFLAVGTNHYLD--- 582
QY 175 DPSSFLREIRTIASDPDERF 194
DB 583 -----AELEISGSKDRTF 596

RESULT 11
Q964N4 ID Q964N4 PRELIMINARY; PRT; 2104 AA.
AC Q964N4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Transmembrane matrix receptor MUP-4.
GN K07D8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

OC Rhabdittidae; Peloderinae; Caenorhabdittis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hong L., Elbl T., Franzini-Armstrong C., Ward J., Rybicka K.K.,  
 RA Gatewood B.K., Bucher E.A.;  
 RT "MUP-4 is a novel matrix receptor with essential functions in  
 RT epithelial cell adhesion at hemidesmosomes and transmission of muscle  
 RT forces in *Caenorhabditis elegans*."  
 RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AF289202; AAK69172.1;  
 DR EMBL; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR000082; SEA domain.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00008; EGF; 16.  
 DR Pfam; PF00092; VWA; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_13.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_3.  
 DR PROSITE; PS01186; EGF\_2; UNKNOWN\_3.  
 DR PROSITE; PS01187; EGF\_CA; UNKNOWN\_2.  
 DR PROSITE; PS50024; SEA; 2.  
 DR PROSITE; PS50234; VMPA; 1.  
 DR RECEPTOR.  
 KW RECEPTOR.  
 SQ SEQUENCE 2104 AA; 230035 MW; 31174C0CCDB61CC2 CRC64;  
 Query Match 18.4%; Score 188.5; DB 5; Length 2104;  
 Best Local Similarity 28.5%; Pred. No. 2.8e-08;  
 Matches 57; Conservative 33; Mismatches 85; Indels 25; Gaps 8;  
 QY 6 VASFPQ---PGSLAPTAQRCPTV--MDVYIVLDGNSIYF--SEVQTFRLRVGLKFI 58  
 DB 411 VVSSNANLPGRVCTLTACPAPQPTDVLFLIDGSSISYFQTEVLAFLAEFFELFPI 470  
 QY 59 DEQIQVGLVGYGSEFVHEMSLGDPTKEEVRAAKNLSRREGRETKTAQAIMVACTEGF 118  
 DB 471 AQOKRVSVVQYSDQIRHFEGLDNGVSDRSLQNALINIIYLGTL-TRTGALEHVANEF 529  
 QY 119 GSHGGRP--EAARLIVVTGESHGDELPAALAKACEGR--VTRYGIAVGHYLROR 174  
 DB 530 SERRGARVPQVSRVAIVITDGSQDNVTRPSD---NARODIQLEFAGVTHNVLVD-- 582  
 QY 175 DPSSFLREIRTIASDPDERF 194  
 DB 583 -----ALDEIISGSKDRTF 596  
 RESULT 12  
 QY3N2 PRELIMINARY; PRT; 526 AA.  
 ID QY3N2;  
 AC QY3N2;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE DJ708FS.1 (Putative novel collagen alpha 1 like protein)  
 DE (Fragment)  
 GN DJ708FS.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smalley C.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL031782; CAB40275.1;  
 DR HSSP; P17301; IAOX.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01391; Collagen; 2.

DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS50234; VMPA; 1.  
 KW Collagen.  
 FT NON TER 526  
 SQ SEQUENCE 526 AA; 57189 MW; 44016B5504F8A5E1 CRC64;  
 Query Match 17.8%; Score 182.5; DB 4; Length 526;  
 Best Local Similarity 30.5%; Pred. No. 1.6e-08;  
 Matches 57; Conservative 36; Mismatches 75; Indels 19; Gaps 6;  
 QY 14 GSLAPTAQRCPTVMDVYIVLDGNSIYF--MSEVQTFRLRVGLKFIPEQIQVGLVYG 71  
 DB 25 GEVRSSCRTAPT--DLVFILDGSYSGPENFEIVKMLVNIITKNFDIPKTIQGVVQYS 82  
 QY 72 ESPVHEMSLGDPTKEEVRAAKNLSRREGRETKTAQAIMVACTEGFSQSHGGRPEAARL 131  
 DB 83 DYPVLEIFPLGSDGSEHILTAVESILYL-GGNTKTGKAIQFALDYLFKAS--SRFLTKI 138  
 QY 132 LVVTGDSHGDGELPAALAKACEGRVTRYGIAVGHYLRQRDPSSFLREIRTIASPD 191  
 DB 139 AVVLITDGKSD--DVKDAQAARDSKITLFAIGV-----GSEIDAEIRAIANKPS 187  
 QY 192 ERPFENV 198  
 DB 188 STYVFTV 194  
 RESULT 13  
 QY3N2 PRELIMINARY; PRT; 954 AA.  
 ID QY3N2;  
 AC QY3N2;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Alpha 1 chain-like collagen COL1A1 precursor.  
 GN COL1A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chou M.-Y.;  
 RT "Cloning and identification of a novel human alpha 1 chain-like  
 RT collagen (COL1A1) gene, a new member of the FACIT family."  
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF330693; AAL50033.1;  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF02210; Collagen; 6.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR ProDom; PD000007; Collagen; 4.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS50234; VMPA; 1.  
 KW Signal; Collagen.  
 FT SIGNAL 1  
 SQ SEQUENCE 954 AA; 99068 MW; C3514675DADCF23 CRC64;  
 Query Match 17.8%; Score 182.5; DB 4; Length 954;  
 Best Local Similarity 30.5%; Pred. No. 3.5e-08;  
 Matches 57; Conservative 36; Mismatches 75; Indels 19; Gaps 6;  
 QY 14 GSLAPTAQRCPTVMDVYIVLDGNSIYF--MSEVQTFRLRVGLKFIPEQIQVGLVYG 71  
 DB 25 GEVRSSCRTAPT--DLVFILDGSYSGPENFEIVKMLVNIITKNFDIGPKTIQGVVQYS 82

```

Qy 72 ESPFHENSIGDPTKEVVRPAANKLSRRRGREKTQAQAIWACTEGFSQSGRGPEEARL 131
Db 83 DYPLVLEPLDSYSGEHLTAVAESILYL-GGNITKGKAIQELADYLFKFS--SRFLTKI 138
Qy 132 LNVVTDGESHDEELPALKACGAGVTRVGIAVTGHYLRGRDPSFLREIRTIASDPD 191
Db 139 AVVLTDSKSD--DVKDAQAARDSKITLFAIGV-----GSETDAELRAIANKPS 187
Qy 192 ERFEEFNV 198
Db 188 STVEFYV 194

```

RESULT 14			
ID	096P44	PRELIMINARY;	PRT; 957 AA.
AC	096P44;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Collagen XXI (Alpha 1 type XXI collagen precursor).		
GN	COL21A1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Fitzgerald J., Bateman J.F.;		
RT	"A new FACIT of the collagen family: COL21A1.";		
RL	FEBS Lett. 0:0-0(2001).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=2185310; PubMed=11863369;		
RA	Chou M.Y., Li H.C.;		
RT	"Genomic Organization and Characterization of the Human Type XXI		
RT	Collagen (COL21A1) Gene.";		
RL	Genomics 79:395-401(2002).		
RL	EMBL; AF14088; AAL02227.1; -		
DR	EMBL; AF438327; AAL86699.1; -		
DR	InterPro: IPR000087; Collagen.		
DR	InterPro: IPR003129; TSPN.		
DR	InterPro: IPR002035; VWF A.		
DR	Pfam; PF01391; Collagen; 6.		
DR	Pfam; PF02210; TSPN; 1.		
DR	Pfam; PFO0092; Vwa; 1.		
DR	ProDom; PD000007; Collagen; 4.		
DR	PROSITE; PS50234; VWA; 1.		
KW	Collagen; Signal.		
FT	SIGNAL 1	22	POTENTIAL.
Q	SEQUENCE 957 AA; 99369 MW; 4AC5CD5E5665A675 CRC64;		

RESULT	15			
ID	Q60863	PRELIMINARY;	PRT;	387 AA.
AC	Q60863;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	von Willebrand factor (Fragment).			
GN	VWF OR VWF.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C;			
RX	MEDLINE=93305993; PubMed=8318738;			
RA	Barrow L.L., Simin K., Mohlke K., Nichols W.C., Ginsburg D.,			
RA	Meisler W.H.;			
RT	"Conserved linkage of neurotrophin-3 and von Willebrand factor on			
RT	mouse chromosome 6.";			
RL	Mamm. Genome 4:343-345.(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C;			
RX	MEDLINE=94250904; PubMed=8193357;			
RA	Nichols W.C., Cooney K.A., Mohlke K.L., Ballew J.D., Yang A.,			
RA	Bruck M.E., Reddington M., Novak E.K., Swank R.T., Ginsburg D.;			
RT	"von Willebrand disease in the RIIIS/J mouse is caused by a defect			
RT	outside of the von Willebrand factor gene [Published erratum appears			
RT	in Blood 1995 Sep 15;86(6):2461].";			
RL	Blood 83:3225-3231.(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C;			
RA	Nichols W.C., Mohlke K.L., Yang A., Bruck M.E., Ginsburg D.;			
RA	Submitted (May-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; U27810; AAA82929.1; -.			
DR	HSSP; P04275; IAUO.			
DR	MCD; MGI:98941; VWF.			
DR	InterPro; IPR002035; VWF_A.			
DR	Pfam; PF00092; vwa; 2.			
DR	SMART; SM00327; VMA; 2.			
DR	PROSITE; PSS0234; VWFA; 2.			
FT	NON_TER 1			
FT	NON_TER 1			
FT	NON_TER 387			
FT	SEQUENCE 387 AA; 43166 MW; D7F8AC5A2465E950 CRC64;			

:

Db                188 STYVFV 194

Search completed: July 16, 2003, 07:55:25  
Job time : 22.3291 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:22 / Search time 22 seconds  
(without alignments)  
1199.256 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025  
Sequence: 1 GICARVDASFOPOGSLAPTA.....FLREIRTIASDPDERFFENV 198

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1025	100.0	1132	21	AAV32243
2	1025	100.0	1152	22	AAAB6457
3	1025	100.0	1152	22	AAAB6458
4	1025	100.0	1167	21	AAV32242
5	1025	100.0	1167	22	AAAB64584
6	912	89.0	195	23	AAU76853
7	912	89.0	195	23	AAU76862
8	797.5	77.8	303	21	AAV32282
9	547	53.4	1188	22	AAAB50087
10	547	53.4	1188	23	AAU10552

11	544	53.1	707	22	AAU19663
12	544	53.1	707	23	ABP47883
13	544	53.1	1034	21	AAAB2590
14	544	53.1	1188	22	AAU14231
15	544	53.1	1188	22	AAU14467
16	544	53.1	1188	22	AAAB30929
17	544	53.1	1188	22	AAAB50085
18	544	53.1	1188	23	AAAB50551
19	544	53.1	1189	21	AAAB25582
20	544	53.1	1189	22	ABG12949
21	492	48.0	193	23	AAU76854
22	492	48.0	193	23	AAU76863
23	467.5	45.6	1179	23	ABAB90759
24	465.5	45.4	1180	23	ABAB90788
25	452.5	44.1	214	22	AAAB50041
26	449.5	43.9	214	22	AAAB50042
27	435	42.4	1367	19	AAW70542
28	430.5	42.0	1183	20	AAU07728
29	419	40.9	1183	20	AAU07728
30	417	40.7	195	23	AAU76851
31	417	40.7	195	23	AAU76860
32	389	38.0	195	23	AAU76852
33	389	38.0	195	23	AAU76861
34	374	36.5	185	22	AAU09125
35	292	28.5	979	22	ABG29239
36	285.5	27.9	148	22	AAU19634
37	285.5	27.9	148	22	AAU19794
38	285.5	27.9	148	23	ABP47854
39	285.5	27.9	148	23	ABP48014
40	282	27.5	103	22	AAU87675
41	282	27.5	103	22	AAU19822
42	282	27.5	103	23	ABP48042
43	244	23.8	176	23	ABB78816
44	209.5	20.4	3063	23	ABAB90762
45	209.5	20.4	3118	22	AAU27790

#### ALIGNMENTS

RESULT 1				
AAV32243	AAV32243 standard; Protein, 1132 AA.			
ID	AAV32243			
XX	AAV32243			
DT	15-FEB-2000 (first entry)			
XX				
DE	Human integrin subunit alpha-10 splice variant.			
XX				
XX	Integrin alpha-10; ISA10; human; trauma; rheumatoid arthritis;			
KW	osteoarthritis; osteoarthritis; cancer; atherosclerosis;			
KW	inflammation; therapy; cartilage; chondrocyte; osteoblast;			
KW	fibroblast; vaccine; marker; splice variant.			
OS	Homo sapiens.			
XX				
EH	Key	Location/Qualifiers		
FT	Peptide	1..22		
FT	Protein	/note= "signal peptide"		
FT		23..1132		
XX		/note= "mature protein"		
PN	WO9951639-A1.			
XX				
PD	14-OCT-1999.			
XX				
PF	31-MAR-1999; 99WO-SE00544.			
PR	02-APR-1998; 98SE-0001164.			
PR	28-JAN-1999; 99SE-0000319.			
XX	(ACT1-) ACTIVE BIOTECH AB.			

Human novel extrac  
Human polypeptide  
Protein encoded by  
Human novel protei  
Human novel protei  
Amino acid sequenc  
Human A259. Homo  
Human A259 polypep  
ITGAI1 protein enc  
Novel human diagn  
Human integrin alp  
Human integrin alp  
Human Tumour Endot  
Rat Tumour Endothe  
Rat alpha1 integr  
Human alpha1 integ  
Integrin alpha-2  
Armenian hamster a  
Human integrin alpe  
Human integrin alp  
Human integrin alp  
Human integrin alp  
Human integrin alp  
Human integrin alp  
Human novel extrac  
Human novel extrac  
Human polypeptide  
Human polypeptide  
Novel central nerv  
Human novel extrac  
Human polypeptide  
Von Willebrand fac  
Human Tumour Endot  
Human full-length

Query Match	Best Local Similarity	100.0%	Score 1025;	DB 21;	Length 1132;
Matches 198;	Conservative	0;	Mismatches	0;	Indels 0;
Gaps	0;				
QY	1	GICARVDSFQFGSLAPTAORCPTFYMDVIVYLDSNSIYPMSEVOTFLRLVGLFLDP	60		
Db	140	GICARVDSFQFGSLAPTAORCPTFYMDVIVYLDSNSIYPMSEVOTFLRLVGLFLDP	199		
QY	61	FOIQVGLVQYGESPVHEWSLGDFTKKEVEVRAAKULSRREGETITQAQIIMVACTGESFQ	120		
Db	200	FOIQVGLVQYGESPVHEWSLGDFTKKEVEVRAAKULSRREGETITQAQIIMVACTGESFQ	259		
QY	121	SHGGPEAPARLLVVVTTDSESHGSELPAALKAACEAGRVTRYIAVLGHYLRQRPSSFL	180		
Db	260	SHGGPEAPARLLVVVTTDSESHGSELPAALKAACEAGRVTRYIAVLGHYLRQRPSSFL	319		
QY	181	REIRTIASDPDERFFNV 198			
Db	320	REIRTIASDPDERFFNV 337			

RESULT 2  
AAB64657  
ID AAB64657 standard; Protein; 1152 AA.

AA64657;  
XX  
XX 22-MAR-2001 (first entry)  
XX  
XX Human secreted protein BLAST search protein SEQ ID NO: 167.

XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KM	antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
KM	vulnerable; anticonvulsant; antibacterial; antitumoral; antiparasitic;
KM	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KM	neurological disease; infection; human; secreted protein.
XX	
OS	Homo sapiens.
XX	
PN	WO20077197-A1.
XX	
PD	21-DEC-2000.
XX	
PF	01-JUN-2000; 2000MO-US14934.
XX	
PR	11-JUN-1999; 98US-0138599.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	(ROSE/) ROSEN C A.
PI	Rosen CA, Ruben SM, Komatsoulis GA;
DR	WPT; 2001-032312/04.
XX	
PT	Isolated nucleic acid molecule encoding a human secreted protein is
XX	used in preventing, treating or ameliorating a medical condition -
PS	Disclosure; Page 543-546; 558pp; English.
CC	The invention relates to the isolation of genes AA32757-F32803 encoding
CC	the human secreted proteins AB64549-B64594. The sequence is a search
CC	result from a BLAST homology search. The genes and proteins are useful
CC	for preventing, ameliorating or treating medical conditions, e.g. by
CC	protein or gene therapy. The genes are isolated from a range of human
CC	tissues disclosed in the specification. The nucleic acids, proteins,
CC	antibodies and (anti)agonists are useful in the diagnosis, treatment
CC	and prevention of: (a) cancer, e.g. breast and ovarian cancer, and
CC	other cancers of the adrenal gland, bone, bone marrow, breast,
CC	gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC	anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC	cardiovascular disorders such as myocardial ischaemias; (d) wound
CC	healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC	and (f) infectious diseases such as viral, bacterial, fungal and
CC	parasitic infections.
XX	
SQ	Sequence 1152 AA;
Query Match	100.0%; Score 1025; DB 22; Length 1152;
Best Local Similarity	100.0%; Pred. No. 7.6e-109;
Matches 198; Conservative	100.0%; Mismatches 0; Indels 0; Gaps 0
DB	1 GICARVDSFPOGSLAETPACRCPYPMVVIVLDGSNSIYPSEVOTFLRLVGKLFIDP 60 140 GICARVDSFPQGSIAETPACRCPYPMVVIVLDGSNSIYPSEVOTFLRLVGKLFIDP 199
OY	61 EOIYGVGLVOYESPYHEWSLGDFTKEEVVRAAKLSRREGGETTQAQIMACTBGEQO 120 200 EOIQVGLVOYESPYHEWSLGDFTKEEVVRAAKLSRREGGETTQAQIMACTBGEQO 259
OY	121 SHGGPEPARLLVVTVDGSHDGEELPALALACEAGRTYRYIAVLGHILRRORDPSSTL 180 260 SHGGPEPARLLVVTVDGSHDGEELPALALACEAGRTYRYIAVLGHILRRORDPSSTL 319
DB	181 REIRTIASDPDERFFNV 198 320 REIRTIASDPDERFFNV 337
OY	
DB	
RESULT 3	
AB64658	
ID	AB64658 standard; Protein; 1152 AA.

```

XX AC AAB64658;
XX DT 22-MAR-2001 (first entry)
XX DE Human secreted protein BLAST search protein SEQ ID NO: 168.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein.
XX OS Homo sapiens.
XX PN WO200071197-A1.
XX PD 21-DEC-2000.
XX PF 01-JUN-2000; 2000WO-US14934.
XX PR 11-JUN-1999; 99US-0138599.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI Rosen CA, Ruben SM, Komatsoulis GA;
XX DR WPI; 2001-032312/04.
XX PT Isolated nucleic acid molecule encoding a human secreted protein is
XX PS used in preventing, treating or ameliorating a medical condition -
XX PS Disclosure; Page 547-551, 558pp; English.
XX CC The invention relates to the isolation of genes AAF32757-F32803 encoding
XX CC the human secreted proteins AAB64549-864594. The sequence is used as a
XX CC query sequence for doing BLASTX searches to identify homologous
XX CC sequences. The genes and proteins are useful for preventing,
XX CC ameliorating or treating medical conditions, e.g. by protein or gene
XX CC therapy. The genes are isolated from a range of human tissues disclosed
XX CC in the specification. The nucleic acids, proteins, antibodies and
XX CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
XX CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
XX CC adrenal gland, bone, bone marrow, breast, gastrointestinal
XX CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
XX CC disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
XX CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular
XX CC disorders such as myocardial ischaemias; (d) wound healing; (e)
XX CC neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX CC infectious diseases such as viral, bacterial, fungal and parasitic
XX CC infections.
XX SQ Sequence 1152 AA;
XX
XX Query Match 100.0%; Score 1025; DB 22; Length 1152;
XX Best Local Similarity 100.0%; Pred. No. 7.6e-109;
XX Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 GICARVDSFPOGSLAPTAQRCPPTVMVIVLDGNSNITYPMSEVOTFLRLVGLFIDP 60
DB 140 GICARVDSFPOGSLAPTAQRCPPTVMVIVLDGNSNITYPMSEVOTFLRLVGLFIDP 199
OY 61 EGIQVGLVOYGSPPHEWLSGPFRTKEEVRAAKULSREGSETTAQIMVACTEGFSQ 120
DB 200 EGIQVGLVOYGSPPHEWLSGPFRTKEEVRAAKULSREGSETTAQIMVACTEGFSQ 259
OY 121 SHGGRPEARLIVVTGSDHGEELPALKACEAGRVRYGIATVGHILRQRDPSSFL 180
DB 260 SHGGRPEARLIVVTGSDHGEELPALKACEAGRVRYGIATVGHILRQRDPSSFL 319
OY 181 REIRTIASDPDERFFENV 198

```

```

DB 320 REIRTIASDPDERFFENV 337
|||||
RESULT 4
ID AAY32242 standard; Protein; 1167 AA.
XX AAY32242;
AC AAY32242;
XX 15-FEB-2000 (first entry)
DT 15-FEB-2000 (first entry)
XX Human integrin subunit alpha-10.
DE Human integrin subunit alpha-10.
XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
XX osteoarthritis; osteoarthritis; cancer; atherosclerosis;
XX inflammation; therapy; cartilage; chondrocyte; osteoblast;
XX fibroblast; vaccine; marker.
XX KW fibroblast; vaccine; marker.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FH 1..22 "signal peptide"
FT /note= "signal peptide"
FT Protein 23..1145
FT /note= "mature protein"
FT Domain 23..1120
FT /note= "extracellular domain"
FT Domain 1121..1145
FT /note= "transmembrane domain"
FT Domain 1122..1167
FT /note= "cytoplasmic domain, specifically claimed
FT in Claim 21"
FT Domain 162..359
FT /note= "I-domain"
FT Binding-site 494..502
FT /note= "cation binding site motif"
FT Binding-site 558..566
FT /note= "cation binding site motif"
FT Binding-site 620..628
FT /note= "cation binding site motif"
FT Modified-site 98
FT /note= "N-glycosylated"
FT Modified-site 336
FT /note= "N-glycosylated"
FT Modified-site 364
FT /note= "N-glycosylated"
FT Modified-site 733
FT /note= "N-glycosylated"
FT Modified-site 839
FT /note= "N-glycosylated"
FT Modified-site 921
FT /note= "N-glycosylated"
FT Modified-site 1018
FT /note= "N-glycosylated"
FT Modified-site 1039
FT /note= "N-glycosylated"
FT Modified-site /note= "N-glycosylated"
XX W09951639-A1.
XX 14-OCT-1999.
XX 31-MAR-1999; 99WO-S000544.
XX 02-APR-1998; 98SE-0001164.
XX 28-JAN-1999; 99SE-0000319.
XX (ACTI-) ACTIVE BIOTECH AB.
XX Lundgren-Akerlund E;
XX WPI; 2000-052639/04.
XX N-PSDB; AA234719.

```

XX New isolated integrin subunit alpha-10, used as a marker or target  
PT molecule for cells during development, regeneration and pathological  
PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
PT inflammation -

XX Claim 1; Fig 6; 90pp; English.

XX This sequence represents novel human chondrocyte integrin subunit  
CC alpha-10 (ISa10). A splice variant is given in AAY32243. The  
CC invention relates to a recombinant or isolated integrin heterodimer  
CC comprising the alpha10 subunit in association with subunit beta  
CC (especially beta-1). The heterodimer and the subunit alpha-10 can  
CC be used as markers or targets of all types of cells, e.g. of  
CC chondrocytes, osteoblasts and fibroblasts. They can also be used:  
CC for treating pathological conditions involving ISa10, such as  
CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
CC for detecting the formation of cartilage during embryonal  
CC development, physiological or therapeutic repair of cartilage,  
CC or detecting regeneration of cartilage or chondrocytes during  
CC transplantation of cartilage or chondrocytes; for selection and  
CC analysis or for sorting, isolating or purification of chondrocytes  
CC and for in vitro studies of differentiation of chondrocytes; and as  
CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
CC skeletal muscle or other tissues where adhesion impairs the function  
CC of the tissue (all claimed). ISa10 binding entities can be used to  
CC determine the differentiation state of cells during embryonic  
CC development, angiogenesis or development of cancer, in pathological  
CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,  
CC in tissue regeneration or in therapeutic and physiological repair  
CC of cartilage (claimed). A vaccine comprising the integrin  
CC heterodimer or subunit alpha-10 is also claimed. ISa10  
CC polynucleotides, vectors, host cells and methods of producing  
CC recombinant ISa10 are also claimed.

XX Sequence 1167 AA;

SQ Query Match 100.0%; Score 1025; DB 21; Length 1167;  
Best Local Similarity 100.0%; Pred. No. 7,7e-109;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GICARVDASFPQGSIAPTAQRCPPTMDVIVLDGNSIYPMSEVQFLRLVGLKFLTDP 60  
DB 140 GICARVDASFPQGSIAPTAQRCPPTMDVIVLDGNSIYPMSEVQFLRLVGLKFLTDP 199  
QY 61 EOIQVGLVOYGESPVHEWSLGDPRTEKEVVAARKNLSREGRRTKTAQAIWVACTEGFSQ 120  
DB 200 EOIQVGLVOYGESPVHEWSLGDPRTEKEVVAARKNLSREGRRTKTAQAIWVACTEGFSQ 259  
QY 121 SHGGRPEARARLLVVTVDGSHDGEELPALAKACEAGRVTGIAVGLHRLRRORDPSSFL 180  
DB 260 SHGGRPEARARLLVVTVDGSHDGEELPALAKACEAGRVTGIAVGLHRLRRORDPSSFL 319  
QY 181 REIRTIASDPDERFFENV 198  
DB -320 REIRTIASDPDERFFENV 337

RESULT 5

AAB64584 standard; Protein; 1167 AA.

XX AAB64584;

XX 22-MAR-2001 (first entry)

XX Human secreted protein #37.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;  
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.

XX OS Homo sapiens.  
XX PN WO20007197-A1.  
XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US14934.

XX 11-JUN-1999; 99US-0138599.

XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM, Komatsoulis GA;

XX WPI: 2001-032312/04.

XX N-PSDB; AAF32793.

XX Isolated nucleic acid molecule encoding a human secreted protein is  
XX used in preventing, treating or ameliorating a medical condition -

XX Claim 11; Page 496-500; 558pp; English.

XX Sequences AAB64549-B64594 represent the amino acid sequences of 47  
CC human secreted proteins encoded by the genes AAF32757-F32803. The genes  
CC and proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
CC a range of human tissues disclosed in the specification. The nucleic  
CC acids, proteins, antibodies and (ant)agonists are useful in the  
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

XX Sequence 1167 AA;

SQ Query Match 100.0%; Score 1025; DB 22; Length 1167;  
Best Local Similarity 100.0%; Pred. No. 7,7e-109;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GICARVDASFPQGSIAPTAQRCPPTMDVIVLDGNSIYPMSEVQFLRLVGLKFLTDP 60  
DB 140 GICARVDASFPQGSIAPTAQRCPPTMDVIVLDGNSIYPMSEVQFLRLVGLKFLTDP 199  
QY 61 EOIQVGLVOYGESPVHEWSLGDPRTEKEVVAARKNLSREGRRTKTAQAIWVACTEGFSQ 120  
DB 200 EOIQVGLVOYGESPVHEWSLGDPRTEKEVVAARKNLSREGRRTKTAQAIWVACTEGFSQ 259  
QY 121 SHGGRPEARARLLVVTVDGSHDGEELPALAKACEAGRVTGIAVGLHRLRRORDPSSFL 180  
DB 260 SHGGRPEARARLLVVTVDGSHDGEELPALAKACEAGRVTGIAVGLHRLRRORDPSSFL 319  
QY 181 REIRTIASDPDERFFENV 198  
DB 320 REIRTIASDPDERFFENV 337

RESULT 6

AAU76853

XX AAU76853 standard; Protein; 195 AA.

XX AAU76853;

XX 21-MAY-2002 (first entry)

XX Human integrin alpha subunit Alpha 10 A domain.

XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;  
 KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
 KW ischaemia-reperfusion injury; immune complex; parasitic disease;  
 KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.  
 XX OS Homo sapiens.  
 XX PN WO200209737-A1.  
 XX PD 07-FEB-2002.  
 XX PF 31-JUL-2001; 2001WO-US23957.  
 XX PR 31-JUL-2000; 2000US-221950P.  
 XX PR 11-JAN-2001; 2001US-0758493.  
 XX PR 13-MAR-2001; 2001US-0805354.  
 XX PA (GEHO) GEN HOSPITAL CORP.  
 XX PI Arnaout AM, Li R, Xiong J;  
 XX DR WPI; 2002-188687/24.  
 XX PT Novel high affinity integrin polypeptide useful for treating restenosis  
 PT and parasitic diseases, comprises all or part of variant integrin alpha  
 PT subunit A domain or variant integrin beta subunit A-like domain  
 XX  
 XX Example 2; Fig 5; 55pp; English.  
 XX PS  
 CC The invention relates to a high affinity integrin polypeptide comprising  
 CC all or part of a variant integrin alpha subunit A domain or a variant  
 CC integrin beta subunit A-like domain. The polypeptide, preferably the  
 CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by  
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
 CC for determining if a test compound is a candidate compound for binding to  
 CC CD1b or for treating an inflammatory disorder, by contacting a test  
 CC compound with the polypeptide and determining if the test compound binds  
 CC to the polypeptide. The integrin subunits are useful for reducing  
 CC skeletal muscle injury, for treating disorders caused by  
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
 CC diseases, to purify variant integrin polypeptide ligands and as bait  
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
 CC the human integrin alpha subunit Alpha 10 A domain.  
 CC  
 XX SQ Sequence 195 AA;  
 Query Match 89.0%; Score 912; DB 23; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-97;  
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 CPTVMDVIVLDGNSIYPMSEVQFLRLVGLKFLPDEQIOVGLVOGESPVHMSLCD 82  
 DB 1 CPTVMDVIVLDGNSIYPMSEVQFLRLVGLKFLPDEQIOVGLVOGESPVHMSLCD 60  
 QY 83 FRTKEVVRRAKNLSRRREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVTGDESHD 142  
 DB 61 FRTKEVVRRAKNLSRRREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVTGDESHD 120  
 QY 143 GELPAAKACAGAVTRYGIAGVILGHYLRORDPSSFLREIRTIASDDPERFFNV 198  
 DB 121 GELPAAKACAGAVTRYGIAGVILGHYLRORDPSSFLREIRTIASDDPERFFNV 176

## RESULT 7

AAU76862  
 ID AAU76862 standard; Protein; 195 AA.

AAU76862;

21-MAY-2002 (first entry)

XX  
 DT  
 XX

DE Human integrin alpha subunit Alpha 10 variant A domain.  
 XX  
 XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;  
 KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
 KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;  
 KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;  
 KW mutein.  
 XX OS Homo sapiens.  
 XX PN Synthetic.  
 XX OS  
 XX FT Key Location/Qualifiers  
 FT Misc-difference 193  
 FT /note= "wild-type Ile substituted by any other amino  
 FT acid"  
 XX  
 XX WO200209737-A1.  
 XX PD 07-FEB-2002.  
 XX PF 31-JUL-2001; 2001WO-US23957.  
 XX PR 31-JUL-2000; 2000US-221950P.  
 XX PR 11-JAN-2001; 2001US-0758493.  
 XX PR 13-MAR-2001; 2001US-0805354.  
 XX PA (GEHO) GEN HOSPITAL CORP.  
 XX PI Arnaout AM, Li R, Xiong J;  
 XX DR WPI; 2002-188687/24.  
 XX PT Novel high affinity integrin polypeptide useful for treating restenosis  
 PT and parasitic diseases, comprises all or part of variant integrin alpha  
 PT subunit A domain or variant integrin beta subunit A-like domain  
 XX  
 XX Claim 53; Page -; 55pp; English.  
 XX PS  
 CC The invention relates to a high affinity integrin polypeptide comprising  
 CC all or part of a variant integrin alpha subunit A domain or a variant  
 CC integrin beta subunit A-like domain. The polypeptide, preferably the  
 CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by  
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
 CC for determining if a test compound is a candidate compound for binding to  
 CC CD1b or for treating an inflammatory disorder, by contacting a test  
 CC compound with the polypeptide and determining if the test compound binds  
 CC to the polypeptide. The integrin subunits are useful for reducing  
 CC skeletal muscle injury, for treating disorders caused by  
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
 CC diseases, to purify variant integrin polypeptide ligands and as bait  
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
 CC a human integrin alpha subunit Alpha 10 variant A domain.  
 CC Note: This variant sequence is not featured in the specification but has  
 CC been derived from the wild-type protein shown in AAU76853.  
 CC  
 XX SQ Sequence 195 AA;  
 Query Match 89.0%; Score 912; DB 23; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-97;  
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 CPTVMDVIVLDGNSIYPMSEVQFLRLVGLKFLPDEQIOVGLVOGESPVHMSLCD 82  
 DB 1 CPTVMDVIVLDGNSIYPMSEVQFLRLVGLKFLPDEQIOVGLVOGESPVHMSLCD 60  
 QY 83 FRTKEVVRRAKNLSRRREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVTGDESHD 142  
 DB 61 FRTKEVVRRAKNLSRRREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVTGDESHD 120  
 QY 143 GELPAAKACAGAVTRYGIAGVILGHYLRORDPSSFLREIRTIASDDPERFFNV 198  
 DB 121 GELPAAKACAGAVTRYGIAGVILGHYLRORDPSSFLREIRTIASDDPERFFNV 176

RESULT 8  
 ID AAY32282  
 AC AAY32282 standard; Protein; 303 AA.  
 XX  
 DT 15-FEB-2000 (first entry)  
 XX  
 DE Mouse integrin subunit alpha-10 partial sequence.  
 XX  
 KW Integrin alpha-10; Itsa10; mouse; trauma; rheumatoid arthritis;  
 osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
 inflammation; therapy; cartilage; chondrocyte; osteoblast;  
 fibroblast; vaccine; marker.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 15  
 FT /note="encoded by TWC"  
 FT  
 XX  
 PN WO9951639-A1.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 31-MAR-1999; 99WO-SE00544.  
 XX  
 PR 02-APR-1998; 98SE-0001164.  
 PR 28-JAN-1999; 99SE-0000319.  
 XX  
 XX  
 PA (ACT1-) ACTIVE BIOTECH AB.  
 XX  
 PI Lundgren-Akerlund E;  
 XX  
 DR WPI; 2000-052639/04.  
 DR N-PSDB; AA234835.  
 XX  
 PT New isolated integrin subunit alpha-10, used as a marker or target  
 PT molecule for cells during development, regeneration and pathological  
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 PT inflammation -  
 XX  
 PS Example 15; Fig 15; 90pp; English.  
 XX  
 CC This sequence represents a polypeptide encoded by an isolated  
 CC partial genomic DNA (see AA234835) of murine integrin subunit  
 CC alpha-10 (Itsa10). The mouse genomic alpha-10 sequence was used  
 CC to generate a targeting vector for knockout experiments. The  
 CC invention relates to a recombinant or isolated integrin heterodimer  
 CC comprising the alpha10 subunit in association with subunit beta  
 CC (especially beta-1). The heterodimer and the subunit alpha-10 can  
 CC be used as markers or targets of all types of cells, e.g. of  
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used  
 CC for treating pathological conditions involving Itsa10, such as:  
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
 CC for detecting the formation of cartilage during embryonal  
 CC development, physiological or therapeutic repair of cartilage,  
 CC or detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes; for selection and  
 CC analysis or for sorting, isolating or purification of chondrocytes  
 CC and for in vitro studies of differentiation of chondrocytes; and as  
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
 CC skeletal muscle or other tissues where adhesion impacts the function  
 CC of the tissue (all claimed). Binding entities, polynucleotides  
 CC (PNS) or oligonucleotides (ONS) can be used for e.g. determining the  
 CC differentiation-state of cells during embryonic development,  
 CC angiogenesis or development of cancer; in pathological conditions  
 CC such as rheumatoid arthritis, osteoarthritis or cancer, in tissue  
 CC regeneration or in therapeutic and physiological repair of  
 CC cartilage (claimed). The PNS and ONS can also be used for detecting  
 CC pathological conditions such as atherosclerosis or inflammation

CC (claimed). A vaccine comprising the integrin heterodimer or  
 CC subunit alpha-10 is also claimed. Vectors, host cells and methods  
 CC of producing recombinant Itsa10 are provided.  
 XX  
 SQ Sequence 303 AA;  
 XX  
 Query Match 77.8%; Score 797.5; DB 21; Length 303;  
 Best Local Similarity 95.1%; Pred. No. 2.2e-83;  
 Matches 156; Conservative 5; Mismatches 2; Indels 1; Gaps 1;  
 XX  
 QY 1 GTCARVDASFPQGSGLATPAGRCPTVMVYVLDGSSNITYPMSEVQTLRLVGLFTDP 60  
 DB 141 GTCARVDASFPQGSGLATPAGRCPTVMVYVLDGSSNITYPMSEVQTLRLVGLFTDP 200  
 QY 61 EGIQVGLVQYGEPPVHEWSLDGPFRTKEVRAAKXLSRREGRETKTAQAIMVACTGEFSQ 120  
 DB 201 EGIQVGLVQYGEPPVHEWSLDGPFRTKEVRAAKXLSRREGRETKTAQAIMVA-TEGFSQ 259  
 QY 121 SHGGRPEARLLVVTDESHDGEELPALAKACEAGRVTRYGIA 164  
 DB 260 SHGGRPEARLLVVTDESHDGEELPALAKACEAGRVTRYGIA 303  
 XX  
 RESULT 9  
 ID AAB50087  
 AC AAB50087 standard; Protein; 1188 AA.  
 XX  
 DT 19-MAR-2001 (first entry)  
 XX  
 DE Murine A259.  
 XX  
 KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;  
 KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;  
 KW rheumatoid arthritis.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT /label= Signal\_peptide  
 FT 23..1188  
 FT /label= Mature\_protein  
 FT 1..1141  
 FT /label= Extracellular\_domain  
 FT 39..74  
 FT /label= Integrin\_alphasubunit\_repeat\_domain\_#1  
 FT 115..157  
 FT /label= Integrin\_alphasubunit\_repeat\_domain\_#2  
 FT 164..345  
 FT /label= I\_domain  
 FT 367..392  
 FT /label= Integrin\_alphasubunit\_repeat\_domain\_#3  
 FT 421..455  
 FT /label= Integrin\_alphasubunit\_repeat\_domain\_#4  
 FT 478..516  
 FT /label= Integrin\_alphasubunit\_repeat\_domain\_#5  
 FT 540..575  
 FT /label= Integrin\_alphasubunit\_repeat\_domain\_#6  
 FT 602..640  
 FT /label= Integrin\_alphasubunit\_repeat\_domain\_#7  
 FT 1142..1164  
 FT /label= Transmembrane\_domain  
 FT 1165..1188  
 FT /label= Cytoplasmic\_domain  
 XX  
 WO200073339-A1.  
 XX  
 PD 07-DEC-2000.  
 XX  
 PF 15-MAY-2000; 2000WO-US13262.  
 XX

PR 28-MAY-1999; 99US-0322790.  
 PR 27-APR-2000; 2000US-0561263.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Pan Y, Lora JM;  
 XX  
 DR WPI; 2001-041142/05.  
 XX N-PSDB; AAC91904, AAC91905.  
 XX  
 PT Nucleic acid encoding alpha-integrin subunits, useful for treatment and  
 PT diagnosis of fibrosis, e.g. of the liver  
 XX  
 PS Claim 8; Fig 5; 164pp; English.  
 XX  
 CC The present sequence is murine integrin alpha subunit, A259. A259 is  
 CC homologous with the alpha and alpha0 integrin subunits and is  
 CC overexpressed in fibrosis. A259 is implicated in regulation of  
 CC proliferation, differentiation and/or function of many different cell  
 CC types. Inhibitors of A259 activity are useful for the treatment of liver  
 CC disease, particularly fibrosis, and also fibrosis in other organs  
 CC (specifically lung and kidney). In addition, A259 can be used for  
 CC treatment and prevention of cancer, osteoporosis, acute myeloid  
 CC leukaemia, HIV infection, and rheumatoid arthritis.  
 XX  
 SQ Sequence 1188 AA;

Query Match 53.4%; Score 547; DB 23; Length 1188;  
 Best Local Similarity 53.0%; Pred. No. 1.5e-53;  
 Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

QY 1 GICARVDASFOPGSLAPTAQRCPTMDVIVILDGNSIYPMSEVQTFRLVGLKFLDP 60  
 DB 137 GICSRVNSNFRFSKTVAPALQRCQTYMDIVILDGNSIYPMVEVQHFLINLKKFYIGP 196  
 QY 61 EGIQVGLVOYGSESPVHMSLGPFRTKEEVYRAKNSRREGTEYTKQAQIMVACTEGESQ 120  
 DB 197 GGIQVGIVOYGEDAHEFLINDYRSYKDVVEAASHIEQRCGETETRTAFGEIPARESEAFQK 256  
 QY 121 SHGSGPEARLILWVTDSHDEELPALKACAGRVTRVGIIVLGHYLRQRQPSSTL 180  
 DB 257 --GGRKGAQKMTVITDGHSDSPDEKVIROSEKDNVTRVAVVAGYNNRGINPEFTL 314  
 QY 181 REIRTIASDPDERFFPNV 198  
 DB 315 NEIKYIASDPDDKHFFNV 332

RESULT 10  
 AAU10552  
 ID AAU10552 standard; Protein; 1188 AA.

AC AAU10552;  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Murine A259 polypeptide.  
 XX  
 KW Mouse; A259; integrin alpha subunit; integrin alpha 10; secreted protein;  
 KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;  
 KW cartilage associated disorder; haematopoietic disorder; bone marrow;  
 KW immune related disease; apoptotic disorder; neuronal tissue disease;  
 KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;  
 KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;  
 KW antidiabetic; antinaeumatic; antiallergic; antidiabetic; dermatological;  
 KW antidiabetic; anticonvulsant; antiparkinsonian.

OS Mus musculus.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..22  
 FT /note= "Signal peptide"  
 FT Domain 1..1141

FT Protein /note= "Extracellular domain"  
 FT 23..1188  
 FT /note= "Mature murine A259"  
 FT 39..74  
 FT Domain /note= "Integrin alpha repeat domain"  
 FT 115..157  
 FT Domain /note= "Integrin alpha repeat domain"  
 FT 164..345  
 FT Domain /note= "I domain or Von Willebrand Factor type A domain"  
 FT 367..392  
 FT Domain /note= "Integrin alpha repeat domain"  
 FT 421..455  
 FT Domain /note= "Integrin alpha repeat domain"  
 FT 478..516  
 FT Domain /note= "Integrin alpha repeat domain"  
 FT 540..575  
 FT Domain /note= "Integrin alpha repeat domain"  
 FT 602..640  
 FT Domain /note= "Integrin alpha repeat domain"  
 FT 1142..1164  
 FT Domain /note= "Transmembrane domain"  
 FT 1165..1188  
 FT Domain /note= "Cytoplasmic domain"

PN W0200181414-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 XX 27-APR-2001; 2001WO-US13516.  
 XX  
 PR 27-APR-2000; 2000US-0561263.

PA (MILL-) MILLENNIUM PHARM INC.

PI Pan Y, Lora J;  
 DR WPI; 2002-041397/05.  
 DR N-PSDB; AAS16874.

PT New A259 nucleic acids and polypeptides, which comprise integrin alpha  
 PT subunit, useful for diagnosing, preventing or treating e.g. liver  
 PT disease, kidney or lung fibrosis, cancers, blood disorders or immune  
 PT related diseases  
 PS Claim 9; Fig 5; 168pp; English.

CC The invention relates to human and murine A259 nucleic acid molecules  
 CC which encode secreted proteins with homology to integrin alpha subunits,  
 CC specifically to integrin alpha 10. The A259 polypeptide and nucleic acid  
 CC are useful for treating liver disease or fibrosis, particularly kidney  
 CC fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also  
 CC useful for diagnosing, preventing or treating cartilage and bone  
 CC associated disorders (such as bone cancer, achondroplasia, myeloma,  
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and  
 CC osteoporosis), bone marrow, blood and haematopoietic disorders (such as  
 CC acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune  
 CC related diseases (such as HIV, viral infections, cancers, T cell  
 CC autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.  
 CC asthma and psoriasis), apoptotic disorders (such as systemic lupus  
 CC erythematosus and insulin-dependent diabetes mellitus), diseases of the  
 CC neuronal tissues (such as epilepsy and muscular dystrophy) and  
 CC neurodegenerative diseases (such as Parkinson's disease and Huntington's  
 CC disease). This sequence represents the murine A259 polypeptide.

SQ Sequence 1188 AA;  
 Query Match 53.4%; Score 547; DB 23; Length 1188;  
 Best Local Similarity 53.0%; Pred. No. 1.5e-53;  
 Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

QY 1 GICARVDASFOPGSLAPTAQRCPTMDVIVILDGNSIYPMSEVQTFRLVGLKFLDP 60  
 DB 137 GICSRVNSNFRFSKTVAPALQRCQTYMDIVILDGNSIYPMVEVQHFLINLKKFYIGP 196





PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465572/50.  
DR N-PEDB; AAS31234.  
XX  
PT Nucleic acid molecules encoding human secreted extracellular matrix  
XX proteins, used in preventing, treating or ameliorating a disorder, e.g.  
PT Alzheimer's and Parkinson's diseases and cancers -  
XX  
PS Claim 11; SEQ ID NO 313; 577pp; English.  
XX  
CC The invention relates to isolated nucleic acid molecules encoding  
CC novel human secreted extracellular matrix proteins (SPs). The  
CC polynucleotides and proteins are used to prevent, treat a medical  
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
CC chickens or sheep. For example, disorders associated with decreased  
CC expression of SPs. The SP polynucleotide or a vector expressing them may  
CC be administered to treat diseases by gene therapy. Antisense molecules  
CC may be administered to down regulate expression of SPs by binding with  
CC the cells own genes and preventing their expression. The polynucleotides  
CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
CC used as antigens to produce antibodies and to identify modulators  
CC (agonists and antagonists) of the SPs. The anti- (SP) antibodies and  
CC antagonists may also be used to down regulate expression and activity of  
CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
CC and multiple sclerosis). Cancers and hyperproliferative disorders (e.g.  
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac  
CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
CC include wound healing, maintenance of organs before transplantation.

Query Match 53.1%; Score 544; DB 22; Length 707;  
Best Local Similarity 53.0%; Pred. No. 1.5e-53;

Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;  
QY 1 GICARVDASFOGSLAFTAQRCPPTMDVIVLVDGNSNIYPMSEVQFLRLRGKLFIDP 60  
DB 153 GMSCRVNSNFRPSKTVAPALRCOTYMDIVLVDGNSNIYPMSEVQFLRLRGKLFIDP 212  
QY 61 EOIQVGLVQGESPVHWSLGDFTKEEVRPAKNLSRESEETKYTAQAINVACTEFSQ 120  
DB 213 GOIQGVVQGEDVDVHEFHLNDYRSVKQVVEAASHIEQRGTEFRTAFGEFARSEAFQX 272  
QY 121 SHGGRPEARLLVVTVDSESHDGEELPALAKACEGRVTRGIAVLGHYLRQDPSSFL 180  
DB 273 --GGRKGAKKVIVITDSESHDSEPLEKVIQOESRDVTRYAVAVLGYNRRGINPETFL 330  
QY 181 REIRTIASDPDERPEFNV 198  
DB 331 NEIKTIASDPDDKGFNV 348  
RESULT 12  
ABP47883  
ID ABP47883 standard; Protein; 707 AA.  
XX  
AC ABP47883;  
XX  
DT 23-AUG-2002 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 313.  
XX  
KM Human, nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KM immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KM antiparkinsonian; antiskinning; antianaemic; antiarthritic; cancer;  
KM antihemetic; hepatotropic; cerebroprotective; antiinflammatory;  
KM antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;  
KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KM neurological disease; infection; neurotropic; gene therapy; vaccine.  
XX  
OS Homo sapiens.  
XX  
XX US2002042386-A1.  
XX  
XX 11-APR-2002.  
XX  
PF 17-JAN-2001; 2001US-0764870.  
XX  
PR 31-JAN-2000; 2000US-179065P.  
PR 04-FEB-2000; 2000US-180628P.  
PR 28-JUN-2000; 2000US-214886P.  
PR 07-JUL-2000; 2000US-216647P.  
PR 07-JUL-2000; 2000US-216880P.  
PR 11-JUL-2000; 2000US-217487P.  
PR 11-JUL-2000; 2000US-217496P.  
PR 14-JUL-2000; 2000US-218290P.  
PR 26-JUL-2000; 2000US-220963P.  
PR 26-JUL-2000; 2000US-220964P.  
PR 14-AUG-2000; 2000US-224518P.  
PR 14-AUG-2000; 2000US-224519P.  
PR 14-AUG-2000; 2000US-225267P.  
PR 14-AUG-2000; 2000US-225268P.  
PR 14-AUG-2000; 2000US-225270P.  
PR 14-AUG-2000; 2000US-225447P.  
PR 14-AUG-2000; 2000US-225757P.  
PR 14-AUG-2000; 2000US-225758P.  
PR 22-AUG-2000; 2000US-226868P.  
PR 30-AUG-2000; 2000US-228924P.  
PR 01-SEP-2000; 2000US-229287P.  
PR 01-SEP-2000; 2000US-229343P.  
PR 01-SEP-2000; 2000US-229344P.  
PR 01-SEP-2000; 2000US-229345P.  
PR 05-SEP-2000; 2000US-229509P.  
PR 05-SEP-2000; 2000US-229513P.  
PR 08-SEP-2000; 2000US-231413P.  
PR 21-SEP-2000; 2000US-234223P.

PR 21-SEP-2000; 2000US-234274P.  
 PR 25-SEP-2000; 2000US-234997P.  
 PR 27-SEP-2000; 2000US-235834P.  
 PR 29-SEP-2000; 2000US-236327P.  
 PR 29-SEP-2000; 2000US-236367P.  
 PR 29-SEP-2000; 2000US-236386P.  
 PR 29-SEP-2000; 2000US-236399P.  
 PR 29-SEP-2000; 2000US-236370P.  
 PR 02-OCT-2000; 2000US-236802P.  
 PR 02-OCT-2000; 2000US-237037P.  
 PR 02-OCT-2000; 2000US-237038P.  
 PR 02-OCT-2000; 2000US-237039P.  
 PR 02-OCT-2000; 2000US-237040P.  
 PR 13-OCT-2000; 2000US-239935P.  
 PR 20-OCT-2000; 2000US-240960P.  
 PR 20-OCT-2000; 2000US-241785P.  
 PR 20-OCT-2000; 2000US-241809P.  
 PR 01-NOV-2000; 2000US-244617P.  
 PR 17-NOV-2000; 2000US-249299P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251869P.  
 XX  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 XX WPI; 2002-470713/50.  
 DR N-PSDB; ABO65558.  
 XX  
 PT New nucleic acid encoding human proteins, useful for diagnosis,  
 PT treatment and prevention of e.g. osteoporosis, also related  
 PT polypeptides and antibodies -  
 XX  
 PS Claim 11; SEQ ID NO 313; 235bp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABO6521-ABO6785) and proteins  
 CC (ABP4746-ABP4810) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer; e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from USPTO at segata.uspto.gov/sequence.html?DocID=999909764870.  
 CC  
 XX  
 XX Sequence 707 AA:  
 Query Match 53.1%; Score 544; DB 23; Length 707;  
 Best Local Similarity 53.0%; Pired. No. 1.5e-53;  
 Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;  
 Oy 1 GICAVDASFOPOGSLAPRACPCPTMDVIVLDGNSNTSPSEVOTFLRLVGLFIDP 60  
 Db 153 GNCSSRNNSFRSKTVAPALQRCQYMDIVILVDSGNSNTSPVVEVQHFINILTKFYIGP 212  
 Oy 61 EOIOGLVGYGSPPHENGLGDFRTKEEVRAKXNLSREGETTAQAQIMVACTEGFSQ 120  
 Db 213 GOIOGVVOYGEDVHEHFLNDYRSKYKVDEAAASHIEQGGTETRTAFIEFAABEAFK 272  
 Oy 121 SHGSPPEARALLVVTGDSHGEELPAALKACENGRVTRYGIAVLGHYLRQRDPSSFL 180

Db 273 --GGRKGAQWIVITDGDSDPDLKEVIOQSERDNTRYAVAVLGYNRRGINPETL 330  
 Oy 181 REIRTIASDPDERFFNV 198  
 Db 331 NEIKYIASDPDKPFNV 348  
 RESULT 13  
 AAB25590  
 ID AAB25590 standard; Protein; 1034 AA.  
 AC AAB25590;  
 DT 21-NOV-2000 (first entry)  
 XX  
 XX Protein encoded by human secreted protein gene #7 clone HOHBV69.  
 XX  
 XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;  
 KW anti-rheumatic; dermatological; antiproliferative; antiarteriosclerotic;  
 KW anticancer; vulnary; antiviral; antibacterial; antifungal;  
 KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;  
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;  
 KW Crohn's disease; nephritis; hyperproliferative disorder;  
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;  
 KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200029435-A1.  
 XX  
 XX 25-MAY-2000.  
 XX  
 XX 27-OCT-1999; 99WO-US25031.  
 XX  
 XX 28-OCT-1998; 98US-0105971.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny J, Moore PA, Wei Y;  
 PI Greene JM;  
 DR WPI; 2000-387742/33.  
 XX  
 PT Isolated nucleic acid molecules encoding human secreted proteins are  
 PT used for the prevention, amelioration and treatment of autoimmune,  
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,  
 PT wounds, and infectious diseases -  
 XX  
 PS Claim 1; Page 678-682; 803bp; English.  
 XX  
 CC The present invention relates to 12 secreted human proteins and the  
 CC nucleotide sequences encoding them. The polynucleotide sequences given  
 CC in AAB80606-A80623 encode the 12 secreted protein sequences given in  
 CC AAB25576-825593. The human secreted proteins have various activities  
 CC dependent on the tissues in which they are expressed. Examples of the  
 CC activities of the proteins include: immunosuppressant;  
 CC anti-inflammatory; antiarthritic; anti-rheumatic; dermatological;  
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnary;  
 CC antiviral; antibacterial; and antifungal activity. The proteins,  
 CC polypeptides, agonists and antagonists may be used to treat prevent  
 CC and/or diagnose various disease, disorders and conditions examples of  
 CC which include: immune disorders e.g. Addison's disease, rheumatoid  
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders  
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;  
 CC hyperproliferative disorders such as paraproteinemia and purpura;  
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;  
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide  
 CC sequences may also be used in wound healing and the treatment of  
 CC infectious diseases. The human secreted protein gene #7 and protein  
 CC sequences are represented in sequences AAB80612 and AAB25582. Secreted  
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences  
 CC AAB80652-A80661 represent genes which are related to the secreted protein  
 CC gene#7.



DR N-PSDB; AAS22772.

PT Isolated polypeptides useful for treating anti-inflammatory diseases.  
PT nervous system disorders, and for regenerating bone and cartilage -

PS Example 4; Page 828-831; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. Polynucleotides of the invention are used as probes and primers, for sequencing, for chromosome or gene mapping, in the production of recombinant proteins, and in generating anti-sense DNA or RNA and in gene therapy. Polypeptides of the invention can be used to target drugs to a tumour, in assays to determine biological activity, to raise antibodies/ elicit an immune response, to determine quantitative protein levels, as tissue markers, and to isolate receptors or ligands. Polypeptides of the invention may also be useful in treating platelet disorders, stem cell disorders, regenerating bone, cartilage, tendon, ligament and/or nerve tissue, wound healing, treating burns, promoting the proliferation, differentiation and survival of stem cells, as a contraceptive, treating osteoporosis and osteoarthritis, anaemia, Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, stroke, immune deficiencies resulting from bacterial, viral or fungal infection or from autoimmunity, cancer, allergy, asthma, graft-versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory diseases, nervous system disorders, and infection. The present sequence represents a protein of the invention.

**SQ Sequence 1188 AA;**

Query Match	53.1%	Score 544	DB 22	Length 1188
Best Local Similarity	53.0%	Pred. No. 3.3e-53		
Matches 105, Conservative	39	Mismatches 52	Indels 2	Gaps 1

[illegible]

Search completed: July 16, 2003, 07:52:05  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 07:55:38 ; Search time 14.318 Seconds  
(without alignments)  
1639.577 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337  
Perfect score: 1025  
Sequence: 1 GICARVDSAPQGSAPTA.....FLREIRTIASDPDERFFENV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues  
Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	912	89.0	195	US-09-805-354-7	Sequence 7, Appli
2	912	89.0	195	US-10-144-259-7	Sequence 7, Appli
3	544	53.1	707	US-10-125-540-313	Sequence 313, App
4	544	53.1	707	US-09-764-870-313	Sequence 313, App
5	544	53.1	1034	US-09-984-130-43	Sequence 43, Appli
6	544	53.1	1034	US-09-836-353A-43	Sequence 43, Appli
7	544	53.1	1189	US-09-984-130-35	Sequence 35, Appli
8	544	53.1	1189	US-09-836-353A-35	Sequence 35, Appli
9	492	48.0	193	US-09-805-354-8	Sequence 8, Appli
10	492	48.0	193	US-10-144-259-8	Sequence 8, Appli
11	467.5	45.6	1151	US-09-984-130-103	Sequence 103, App
12	467.5	45.6	1151	US-09-836-353A-103	Sequence 103, App
13	452.5	44.1	214	US-10-061-658-5	Sequence 5, Appli
14	449.5	43.9	214	US-10-061-658-6	Sequence 6, Appli
15	449.5	43.9	214	US-10-061-658-9	Sequence 9, Appli
16	441	43.0	1181	US-10-160-354-2	Sequence 2, Appli
17	433.5	42.3	212	US-09-996-738-5	Sequence 5, Appli
18	430.5	42.0	212	US-09-996-738-6	Sequence 6, Appli
19	427	41.7	195	US-09-805-354-5	Sequence 5, Appli

20	427	41.7	195	US-10-144-259-5	Sequence 5, Appli
21	399	38.9	195	US-09-805-354-6	Sequence 6, Appli
22	399	38.9	195	US-10-144-259-6	Sequence 6, Appli
23	285.5	27.9	148	US-10-125-540-284	Sequence 284, App
24	285.5	27.9	148	US-10-125-540-444	Sequence 444, App
25	285.5	27.9	148	US-09-764-870-284	Sequence 284, App
26	285.5	27.9	148	US-09-764-870-444	Sequence 444, App
27	282	27.5	103	US-10-125-540-472	Sequence 472, App
28	282	27.5	103	US-09-764-870-472	Sequence 472, App
29	210.5	20.5	3063	US-10-177-293-63	Sequence 63, Appli
30	209.5	20.4	3063	US-10-177-293-61	Sequence 61, Appli
31	198	19.3	413	US-09-891-943-101	Sequence 101, App
32	198	19.3	413	US-09-350-259-101	Sequence 101, App
33	191	18.6	776	US-10-000-512-8	Sequence 8, Appli
34	191	18.6	915	US-09-905-291A-34	Sequence 34, Appli
35	191	18.6	915	US-09-902-853-34	Sequence 34, Appli
36	191	18.6	915	US-09-907-824-34	Sequence 34, Appli
37	191	18.6	915	US-09-907-841-34	Sequence 34, Appli
38	191	18.6	915	US-09-904-011-34	Sequence 34, Appli
39	191	18.6	915	US-10-028-072-294	Sequence 294, App
40	191	18.6	915	US-09-906-742-34	Sequence 34, Appli
41	191	18.6	915	US-10-121-049-294	Sequence 294, App
42	191	18.6	915	US-10-123-904-294	Sequence 294, App
43	191	18.6	915	US-10-140-470-294	Sequence 294, App
44	191	18.6	915	US-09-906-838-34	Sequence 34, Appli
45	191	18.6	915	US-09-907-613-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1  
US-09-805-354-7  
Sequence 7, Application US/09805354  
Publication No. US2003078375A1  
GENERAL INFORMATION:  
APPLICANT: Arnaout, M. Amin  
APPLICANT: Li, Rui  
TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF  
FILE REFERENCE: 00786-536001  
CURRENT APPLICATION NUMBER: US/09/805,354  
CURRENT FILING DATE: 2002-06-04  
PRIOR APPLICATION NUMBER: US 09/758,493  
PRIOR FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: US 60/221,950  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 195  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-805-354-7

Query Match 89.0%; Score 912; DB 9; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1e-89;  
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 CPTVMDVIVDGSNSIYPMSEVOTFLRLVYKLFIDPEQIQVGLVYGESPVHEWSIGCD 82  
DB 1 CPTVMDVIVDGSNSIYPMSEVOTFLRLVYKLFIDPEQIQVGLVYGESPVHEWSIGCD 60  
QY 83 FRTREVVRAKNLSRRRGRETAKQAQIMVACTGFSQSHGGRPEAAALLVVPTDGESHD 142  
DB 61 FRTREVVRAKNLSRRRGRETAKQAQIMVACTGFSQSHGGRPEAAALLVVPTDGESHD 120  
QY 143 GEEUPAALKACAGVTRVYGAIVLGHYLRORDSSFLREIRTIASDPDERFFENV 198  
DB 121 GEEUPAALKACAGVTRVYGAIVLGHYLRORDSSFLREIRTIASDPDERFFENV 176

RESULT 2



US-09-984-130-43

Query Match	53.1%;	Score 544;	DB 9;	Length 1034;
Best Local Similarity	53.0%;	Pred. No. 2.7e-49;		
Matches 105;	Conservative 39;	Mismatches 52;	Indels 2;	Gaps 1

```
QY      GI CARVDASPOFGSLAPTAORCEPTMDVIVILDSNSITYPESEVOGTRELTVKFLFDP 60
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      GMSGRVSNRSREKXTA PALQRCTIMDIVILDSNSITYPWEVOHFLINILKKFYTG 196
       137

QY      EOIQVGIVQYESPVHEWISLGPRTKEEVRAAKULSRREGRETQAIAINWACTEESFQ 120
       61
Db      GQIQGVGVQGEDVDVNEHPNDVRYSKVDAASHIEDRGGETRTAGIEFANSEAFOX 256
       197

QY      SHGGRPEAAALVVVTDDGESHDGEELPALKAACEGATRYGIAVLGHYLARODPSFPL 180
       121
Db      --GGRKAKAVMIVILDGESHS DPLEKVIQOESERDNTRYAVAVALGYNNRGINPETFL 314
       257

QY      REIRTIASDPDEREFNFV 198
       181
Db      NBIKIYASDPDDDKHFNV 332
       315
```

RESULT 6  
IIS-09-83

```

Sequence 43, Application US/09836353A
Publication No. US20030129685A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: PF489P1
CURRENT APPLICATION NUMBER: US/09/836,353A
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 147
SOFTWARE: Patentn Ver. 2.0
SEQ ID NO 43
LENGTH: 1034
TYPE: PRT
ORGANISM: Homo sapiens
US-09-836-353A-43

```

Query Match	53.1%;	Score 544;	DB 9;	Length 1034;
Best Local Similarity	53.0%;	Pred. No. 2.7e-49;		
Matches 105;	Conservative 39;	Mismatches 52;	Indels 2;	Gaps 1

[illegible]

```

RESULT 7
US-09-984-130-35
; Sequence 35, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/199,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-35

Query Match          53.1%   Score 544;   DB 9;   Length 1189;
Best Local Similarity 53.0%;   Pred. NO. 3; 3e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

```

	Query March	53.1%;	Score 544;	DB 9;	Length 1189;
	Best Local Similarity	53.0%;	Pred. No. 3.3e-49;		
	Matches	105;	Conservative 39;	Mismatches 52;	Indels 2; Gaps 1;
Oy	1	GICARVDASFOGQSLAPTAQRCPFTYMDVIVLDGSSNIYSFWEVQFLRLRVGKLFIDP	60		
Db	137	GMCSRNVSNFNFSKTVPALORCCTYMDIVILDGSSNIYSFWVVOHFLINLKKFYGP	196		
Oy	61	EQIOVGAVOYGESPVHMESLCDPRTKEEVPAANKLSRSEGEREKTQAQIIVACBEGSQ	120		
Db	197	GQIQGVAVQJEDDVYHEFLNDYNSVKDVEAASHIEGRGTERTIANGIFAFASEAQX	256		
Oy	121	SHGGRPEARLLVVVTLDGESHDGEELPALAKACEAGRYTRYGI AVLGHYLRRORDPSFFL	180		
Db	257	--GGRKAQKAMVIITGDESHSDPLEKVIQQSERDNVTRVAVAVLGYNNRGINPETFL	314		
Oy	181	REIRTIASDDPEREFENV	198		
Db	315	NEIKYIASDPDKHFENV	332		

```

1      RESULT 8
2      US-09-836-353A-35
3      : Sequence 35, Application US/09836353A
4      : Publication No. US20030129665A1
5      :
6      : GENERAL INFORMATION:
7      :
8      :   APPLICATION: N1 et al.
9      :   TITLE OF INVENTION: 12 Human Secreted Proteins
10     :   FILE REFERENCE: PFA69P1
11     :   CURRENT APPLICATION NUMBER: US/09/836,353A
12     :   CURRENT FILING DATE: 2001-04-18
13     :   PRIOR APPLICATION NUMBER: 60/198,407
14     :   PRIOR FILING DATE: 2000-04-19
15     :   PRIOR APPLICATION NUMBER: PCT/US99/25031
16     :   PRIOR FILING DATE: 1999-10-27
17     :   PRIOR APPLICATION NUMBER: 60/105,971
18     :   PRIOR FILING DATE: 1998-10-28
19     :   NUMBER OF SEQ ID NOS: 147
20     :   SOFTWARE: PatentIn Ver. 2.0
21     :
22     :   SEQ ID NO: 35
23     :   LENGTH: 1169
24     :   TYPE: PRT
25     :   ORGANISM: Homo sapiens
26     :
27     : US-09-836-353A-35

```

Query Match	53.1%	Score 544	DB 9	Length 1189
Best Local Similarity	53.0%	Pred. No. 3.3e-49		
Matches 105	Conservative 39	Mismatches 52	Indels 2	Gaps 1
QY	1	GTCARVVASFPQSLAPTAQRCPTDVIIVLDGNSISITPWSVQTFRLRVKLFIDP	60	

```

; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-259-8
Db 137 GMSRSVNSNFRFSKTVAPALQRCQTYMDIYIVLDGNSIYPMWEVQHFLINILKKFYICP 196
QY 61 EQIQVGLVOYGESPVHEWSLGDFTKKEVVRRAKNLSRREGRTKTAQAIVACTGFSQ 120
Db 197 GQIQVGVVOYGEDVHEFLINDYRSYKDVVEASHLEQRCGTETRTAFGIEFARSEAFQK 256
QY 121 SHGGRPEARLLVVTVDGESHDEELPAALKACAGAVTRYGIAVLGHYLRORDPSSFL 180
Db 257 --GGRGAKKVMIVITDGSHEHSDPLEKVIQOSEBDNVTYVAVALGYINRKGINPFTL 314
QY 181 REIRTIASDPDERFFENV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 9
US-09-805-354-8
; Sequence 8, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaud, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805,354
; PRIOR FILING DATE: 2002-06-04
; PRIOR FILING DATE: 2001-01-11
; PRIOR FILING DATE: 2001-01-11
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-8
Query Match 48.0%; Score 492; DB 9; Length 193;
Best Local Similarity 54.5%; Pred. No. 9, 8e-45;
Matches 96; Conservative 31; Mismatches 47; Indels 2; Gaps 1;

QY 23 CPTVMDVIVLDGNSIYPMWEVQTFRLRVGKLFIDPQIQVGLVOYGESPVHEWSLGD 82
Db 1 CQTYMDIYIVLDGNSIYPMWEVQHFLINILKKFYIGQIQVGVVOYGEDVHEFLIND 60
QY 83 FRTKEEVRAAKNLSRREGRTKTAQAIVACTGFSQSHGGRPEARLLVVTVDGESH 142
Db 61 YRSVKDVVEASHLEQRCGTETRTAFGIEFARSEAFQK--GGRGAKKVMIVITDGS 118
QY 143 GEELPAALKACAGAVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 198
Db 119 SPDEKVIQOSEBDNVTYVAVALGYINRKGINPFTLEIKYIASDPDDKHFFNV 174

RESULT 10
US-10-144-259-8
; Sequence 8, Application US/10144259
; Publication No. US20030109691A1
; GENERAL INFORMATION:
; APPLICANT: Arnaud, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-548001
; CURRENT APPLICATION NUMBER: US/10/144,259
; CURRENT FILING DATE: 2002-09-04
; PRIOR FILING DATE: 2001-01-11
; PRIOR FILING DATE: 2001-01-11
; PRIOR FILING DATE: 2000-07-31
; PRIOR FILING DATE: 2000-07-31
```

```

; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-259-8
Query Match 48.0%; Score 492; DB 9; Length 193;
Best Local Similarity 54.5%; Pred. No. 9, 8e-45;
Matches 96; Conservative 31; Mismatches 47; Indels 2; Gaps 1;

QY 23 CPTVMDVIVLDGNSIYPMWEVQTFRLRVGKLFIDPQIQVGLVOYGESPVHEWSLGD 82
Db 1 CQTYMDIYIVLDGNSIYPMWEVQHFLINILKKFYIGQIQVGVVOYGEDVHEFLIND 60
QY 83 FRTKEEVRAAKNLSRREGRTKTAQAIVACTGFSQSHGGRPEARLLVVTVDGESH 142
Db 61 YRSVKDVVEASHLEQRCGTETRTAFGIEFARSEAFQK--GGRGAKKVMIVITDGS 118
QY 143 GEELPAALKACAGAVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 198
Db 119 SPDEKVIQOSEBDNVTYVAVALGYINRKGINPFTLEIKYIASDPDDKHFFNV 174

RESULT 11
US-09-984-130-103
; Sequence 103, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; PRIOR FILING DATE: 2001-10-29
; PRIOR FILING DATE: 2000-10-30
; PRIOR FILING DATE: 2000-10-30
; PRIOR FILING DATE: 1999-10-27
; PRIOR FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2000-04-19
; PRIOR FILING DATE: 2000-04-19
; PRIOR FILING DATE: 2000-04-19
; PRIOR FILING DATE: 1999-10-27
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-103
Query Match 45.6%; Score 467.5; DB 9; Length 1151;
Best Local Similarity 46.5%; Pred. No. 4, 9e-41;
Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

QY 1 GICARVDSAPQOSIAPTACRCPTVMDVIVLDGNSIYPMWEVQTFRLRVGKLFIDP 60
Db 118 GICSDVSPFTFOYVANSIAP-VEGCTQIDIVIVLDGNSIYPMWDSVTALNDLKRMDIGP 176
QY 61 EQIQVGLVOYGESPVHEWSLGDFTKKEVVRRAKNLSRREGRTKTAQAIVACTGFSQ 120
Db 177 KQIQVGIQYGENVTHEENLNKYSTEVVLAAKIYVGRGROMTALGTGTAKKAEATE 236
QY 121 SHGGRPEARLLVVTVDGESHDEELPAALKACAGAVTRYGIAVLGHYLRORDPSSFL 180
Db 237 ARGARRGVKVMIVITDGSHEHSDHNRKLVQIDCEDENIQFSIALGSGYNRGNLSTKTFV 296
QY 181 REIRTIASDPDERFFENV 198
Db 297 EIKSIASEPTEKHFENV 314
```



RESULT 12  
US-09-836-353A-103  
; Sequence 103, Application US/09836353A  
; Publication No. US20030129685A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: PF489P1  
; CURRENT APPLICATION NUMBER: US/09/836,353A  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,407  
; PRIOR FILING DATE: 2000-04-15  
; PRIOR APPLICATION NUMBER: PCT/US99/25031  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR APPLICATION NUMBER: 60/105,971  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 103  
; LENGTH: 1151  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-836-353A-103

Query Match 45.6%; Score 467.5; DB 9; Length 1151;  
Best Local Similarity 46.5%; Pred. No. 4.9e-41;  
Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

QY 1 GICAVDASFPQGS LAPAORCPITMDVIVYLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQV 60  
DB 118 GICSVSPFPQVNSIAP-VQECSTQLDIVYLDGNSIYPMDSVTAFINDLKRMDIGP 176  
QY 61 EIOVGLVOYGSPPVHEMSIGDPRFTRKEEVVRAAKNLSRREGRETKQAQIMVACTEGFSQ 120  
DB 177 KOTQVIGVQYGENVTHEFNLKYSSTEEVLVAKKIVQRGQOTMTALCTDTRKAEATE 236  
QY 121 SHGPEEARLLVVTDSGSHDGEELPALKACEAGRVTRYGIAVGHYLRQRDPSSFL 180  
DB 237 ARGARGVYKVVIVTDSGSHDNHRLKAVIQDCEDENIQRFSAIIGSYNRGNLSTEKVEV 296  
QY 181 REIRTIASDPDERFPFNV 198  
DB 297 EIKSIASEPTEKHFFNV 314

RESULT 13  
US-10-061-658-5  
; Sequence 5, Application US/10061658  
; Publication No. US20020182213A1  
; GENERAL INFORMATION:  
; APPLICANT: Biogen, Inc.  
; APPLICANT: Gotwals, Philip  
; APPLICANT: Koteliansky, Victor  
; TITLE OF INVENTION: Method for the Treatment of Fibrosis  
; FILE REFERENCE: A073US  
; CURRENT APPLICATION NUMBER: US/10/061,658  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 60/137,214  
; PRIOR FILING DATE: 1999-06-01  
; PRIOR APPLICATION NUMBER: 60/130,847  
; PRIOR FILING DATE: 1999-04-22  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Rat  
US-10-061-658-5

Query Match 44.1%; Score 457.5; DB 9; Length 214;  
Best Local Similarity 46.6%; Pred. No. 1.9e-40;  
Matches 90; Conservative 34; Mismatches 68; Indels 1; Gaps 1;

QY 6 VDASFQGS LAPAORCPITMDVIVYLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQV 65  
DB 1 VSPFQVNSIAP-VQECSTQLDIVYLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQV 59  
QY 66 GLVOYGSPPVHEMSIGDPRFTRKEEVVRAAKNLSRREGRETKQAQIMVACTEGFSQSHGR 125  
DB 60 GIOVYGENVTHEFNLKYSSTEEVLVAKKIVQRGQOTMTALCTDTRKAEATEARGAR 119  
QY 126 PEARLLVVTDSGSHDGEELPALKACEAGRVTRYGIAVGHYLRQRDPSSFLREIRT 185  
DB 120 RGKVKWIVTDSGSHDNHRLKAVIQDCEDENIQRFSAIIGSYNRGNLSTEKVEEIKS 179  
QY 186 IASDPDERFPFNV 198  
DB 180 IASEPTEKHFFNV 192

RESULT 14  
US-10-061-658-6  
; Sequence 6, Application US/10061658  
; Publication No. US20020182213A1  
; GENERAL INFORMATION:  
; APPLICANT: Biogen, Inc.  
; APPLICANT: Gotwals, Philip  
; APPLICANT: Koteliansky, Victor  
; TITLE OF INVENTION: Method for the Treatment of Fibrosis  
; FILE REFERENCE: A073US  
; CURRENT APPLICATION NUMBER: US/10/061,658  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 60/137,214  
; PRIOR FILING DATE: 1999-06-01  
; PRIOR APPLICATION NUMBER: 60/130,847  
; PRIOR FILING DATE: 1999-04-22  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-061-658-6

Query Match 43.9%; Score 449.5; DB 9; Length 214;  
Best Local Similarity 46.1%; Pred. No. 4e-40;  
Matches 89; Conservative 35; Mismatches 68; Indels 1; Gaps 1;

QY 6 VDASFQGS LAPAORCPITMDVIVYLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQV 65  
DB 1 VSPFQVNSIAP-VQECSTQLDIVYLDGNSIYPMDSVTAFINDLKRMDIGPKQTV 59  
QY 66 GLVOYGSPPVHEMSIGDPRFTRKEEVVRAAKNLSRREGRETKQAQIMVACTEGFSQSHGR 125  
DB 60 GIOVYGENVTHEFNLKYSSTEEVLVAKKIVQRGQOTMTALCTDTRKAEATEARGAR 119  
QY 126 PEARLLVVTDSGSHDGEELPALKACEAGRVTRYGIAVGHYLRQRDPSSFLREIRT 185  
DB 120 RGKVKWIVTDSGSHDNHRLKAVIQDCEDENIQRFSAIIGSYNRGNLSTEKVEEIKS 179  
QY 186 IASDPDERFPFNV 198  
DB 180 IASEPTEKHFFNV 192

RESULT 15  
US-10-061-658-9  
; Sequence 9, Application US/10061658  
; Publication No. US20020182213A1  
; GENERAL INFORMATION:  
; APPLICANT: Biogen, Inc.  
; APPLICANT: Gotwals, Philip  
; APPLICANT: Koteliansky, Victor  
; TITLE OF INVENTION: Method for the Treatment of Fibrosis  
; FILE REFERENCE: A073US



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:49:17 ; Search time 7.79747 Seconds  
(without alignments)  
747.132 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025  
Sequence: 1 GICARVDASFOQGLAPTA.....FLREIRTIASDPDERFFPNV 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfltest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	205	20.0	496	1 US-08-462-128-37	Sequence 37, Appl
2	205	20.0	496	1 US-08-463-180-37	Sequence 37, Appl
3	205	20.0	496	2 US-08-001-078A-1	Sequence 1, Appl
4	205	20.0	496	2 US-08-897-443-4	Sequence 4, Appl
5	205	20.0	496	2 US-08-463-218-1	Sequence 1, Appl
6	205	20.0	496	5 PCT-US94-00253-1	Sequence 1, Appl
7	198	19.3	413	1 US-08-485-618-101	Sequence 101, App
8	198	19.3	413	2 US-08-605-672-101	Sequence 101, App
9	198	19.3	413	2 US-08-482-293A-101	Sequence 101, App
10	198	19.3	413	2 US-08-943-363-101	Sequence 101, App
11	198	19.3	413	4 US-09-193-043-101	Sequence 101, App
12	198	19.3	413	4 US-09-688-307A-101	Sequence 101, App
13	196	19.1	956	2 US-08-897-443-3	Sequence 3, Appl
14	191	18.6	638	2 US-08-897-443-1	Sequence 3, Appl
15	177.5	17.3	1178	1 US-08-199-776-2	Sequence 2, Appl
16	177.5	17.3	1178	3 US-08-663-731-2	Sequence 2, Appl
17	177.5	17.3	1178	3 US-08-879-338-2	Sequence 2, Appl
18	177.5	17.3	1178	5 PCT-US95-02044-2	Sequence 2, Appl
19	177.5	17.3	1179	4 US-09-293-238B-2	Sequence 2, Appl
20	169	16.5	1155	1 US-08-286-889-46	Sequence 46, Appl
21	169	16.5	1155	1 US-08-485-618-46	Sequence 46, Appl
22	169	16.5	1155	1 US-08-362-652-46	Sequence 46, Appl
23	169	16.5	1155	2 US-08-605-672-46	Sequence 46, Appl
24	169	16.5	1155	2 US-08-482-293A-46	Sequence 46, Appl
25	169	16.5	1155	2 US-08-943-363-46	Sequence 46, Appl
26	169	16.5	1155	4 US-09-193-043-46	Sequence 46, Appl
27	169	16.5	1155	4 US-09-688-307A-46	Sequence 46, Appl

28	169	16.5	1161	1 US-08-485-618-53	Sequence 53, Appl
29	169	16.5	1161	2 US-08-362-652-53	Sequence 53, Appl
30	169	16.5	1161	2 US-08-605-672-53	Sequence 53, Appl
31	169	16.5	1161	2 US-08-482-293A-53	Sequence 53, Appl
32	169	16.5	1161	2 US-08-943-363-53	Sequence 53, Appl
33	169	16.5	1161	4 US-09-193-043-53	Sequence 53, Appl
34	169	16.5	1161	4 US-09-688-307A-53	Sequence 53, Appl
35	164	16.0	2813	3 US-08-485-618-93	Sequence 93, Appl
36	164	16.0	2813	3 US-09-132-652-2	Sequence 2, Appl
37	161	15.7	264	1 US-08-485-618-93	Sequence 93, Appl
38	161	15.7	264	1 US-08-362-652-93	Sequence 93, Appl
39	161	15.7	264	2 US-08-605-672-93	Sequence 93, Appl
40	161	15.7	264	2 US-08-482-293A-93	Sequence 93, Appl
41	161	15.7	264	2 US-08-943-363-93	Sequence 93, Appl
42	161	15.7	264	4 US-09-193-043-93	Sequence 93, Appl
43	161	15.7	264	4 US-09-688-307A-93	Sequence 93, Appl
44	160	15.6	1161	1 US-08-172-497-2	Sequence 2, Appl
45	160	15.6	1161	1 US-08-286-889-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-462-128-37  
; Sequence 37, Application US/08462128  
; Patent No. 5686059  
; GENERAL INFORMATION:  
; APPLICANT: Goetlinck, Paul F.  
; APPLICANT: Tondra, David, Mehridad  
; TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR  
; TITLE OF INVENTION: USE  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lohive & Cockfield  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,128  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/006,096  
; FILING DATE: 15-JAN-1993  
; APPLICATION NUMBER: US 07/866,403  
; FILING DATE: 10-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Paul L.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: MGP-005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 496 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-462-128-37  
; Query Match 20.0%; Score 205; DB 1; Length 496;  
; Best Local Similarity 33.5%; Pred. NO. 4.4e-17;  
; Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;  
; QY 28 DVIIVLDGNSIYR--WSEVQTFRLRLVGLKFLIDPEIQVGLVQYGSPEVHWSLGDPRRT 85

Db 275 DLVFLIDGSKSVRPENFELVKRFTSQIIVTLVDSDLAQVGVSSVRQEPFLGRFHT 334  
QY 86 KEEVRAAKNLRRREGRETKTAQAIWVACTEGFSQSHGCRPEARLLVVTGDSHGDEE 145  
Db 335 KKDIAKAVNNMSYME-KGTMGALKYLLDINSFTVSSGARPGAKVGIVFTDGRSODYIN 393  
QY 146 LPAALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLEIRITIASDP-DERFF 196  
Db 394 -DAKAKDLGFKMFVGVG-----NAVEDELRITASEPVAEHYFY 433

RESULT 2  
US-08-463-180-37  
Sequence 37, Application US/08463180  
Patent No. 5741670  
GENERAL INFORMATION:  
APPLICANT: Goeltnck, Paul F.  
APPLICANT: Tondra, M., Mehrdad  
TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lahive & Cockfield  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,180  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/006,096  
FILING DATE: 15-JAN-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/866,403  
FILING DATE: 10-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: MGP-005DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 496 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-463-180-37

Query Match 20.0%; Score 205; DB 1; Length 496;  
Best Local Similarity 33.5%; Pred. No. 4.4e-17;  
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

QY 28 DVIYVLDGNSIYF-WSEVOTFLRLVYKLFIDPEQIQVGLVQGESPVHWSIGDFRT 85  
Db 275 DLVFLIDGSKSVRPENFELVKRFTSQIIVTLVDSDLAQVGVSSVRQEPFLGRFHT 334  
QY 86 KEEVRAAKNLRRREGRETKTAQAIWVACTEGFSQSHGCRPEARLLVVTGDSHGDEE 145  
Db 335 KKDIAKAVNNMSYME-KGTMGALKYLLDINSFTVSSGARPGAKVGIVFTDGRSODYIN 393  
QY 146 LPAALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLEIRITIASDP-DERFF 196

Db 394 -DAKAKDLGFKMFVGVG-----NAVEDELRITASEPVAEHYFY 433

RESULT 3  
US-08-001-078A-1  
Sequence 1, Application US/08001078A  
Patent No. 5872094  
GENERAL INFORMATION:  
APPLICANT: Goeltnck, Paul F.  
APPLICANT: Tondra, M., Mehrdad  
TITLE OF INVENTION: METHODS FOR PROMOTING CARTILAGE MATRIX  
FORMATION  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 STATE STREET, Suite 510  
CITY: BOSTON  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/001,078A  
FILING DATE: 06-JAN-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: MGP-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 496 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-001-078A-1

Query Match 20.0%; Score 205; DB 2; Length 496;  
Best Local Similarity 33.5%; Pred. No. 4.4e-17;  
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

QY 28 DVIYVLDGNSIYF-WSEVOTFLRLVYKLFIDPEQIQVGLVQGESPVHWSIGDFRT 85  
Db 275 DLVFLIDGSKSVRPENFELVKRFTSQIIVTLVDSDLAQVGVSSVRQEPFLGRFHT 334  
QY 86 KEEVRAAKNLRRREGRETKTAQAIWVACTEGFSQSHGCRPEARLLVVTGDSHGDEE 145  
Db 335 KKDIAKAVNNMSYME-KGTMGALKYLLDINSFTVSSGARPGAKVGIVFTDGRSODYIN 393  
QY 146 LPAALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLEIRITIASDP-DERFF 196  
Db 394 -DAKAKDLGFKMFVGVG-----NAVEDELRITASEPVAEHYFY 433

RESULT 4  
US-08-897-443-4  
Sequence 4, Application US/08897443  
Patent No. 5981263  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
APPLICANT: Kaser, Mathew  
TITLE OF INVENTION: HUMAN MATRILIN-3



```

QY 28 DVVLVDLGSNSIIP--MSEVQIFLRRLVYKFLIDBEQVLYGVGVGSEPVHNMSTLGPRT 85
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 275 DLVFLIDGSSKVPENFELVKFISQIVDTLLVSDKLAQVGLVQVSSSRVQEGFPLGRFHT 334
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 86 KEVVRRAKKLSRREGETKTAQIIVACTEGSGSHGRPEAAFLVWVVDGESHGEE 145
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 335 KKDICKAALVRMSTWYE-RGTVTGAALKYLLIDNSTVSSGRPEAQVGLVFTDGRSDIYN 393
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 146 LPALAKACEAG-RVTRVGLAVLGHVLRDRDPSSFLEIRTIATSPD-DEPF 196
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 394 -DAAKKAKDGFPMFAVGVG-----NAVEDELRELTASPEVAHEHYF 433
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

1      RESULT 7
2      US-08-485-618-101
3      Sequence 101, Application US/08485618
4      Patent No. 5728533
5      GENERAL INFORMATION:
6      APPLICANT: Gallatin, W. Michael
7      APPLICANT: Van der Vieren, Monica
8      TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
9      NUMBER OF SEQUENCES: 103
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
12     STREET: 233 South Wacker Drive, 6300 Sear Tower
13     CITY: Chicago
14     STATE: Illinois
15     COUNTRY: United States
16     ZIP: 60606-6402
17     COMPUTER READABLE FORM:
18     MEDIUM TYPE: Floppy disk
19     COMPUTER: IBM PC compatible
20     OPERATING SYSTEM: PC-DOS/MS-DOS
21     SOFTWARE: PatentIn Release #1.0, Version #1.25
22     CURRENT APPLICATION DATA:
23     APPLICATION NUMBER: US/08/485,618
24     FILING DATE:
25     CLASSIFICATION: 435
26     PRIOR APPLICATION DATA:
27     APPLICATION NUMBER: US 08/173,497
28     FILING DATE: 23-DEC-1993
29     PRIOR APPLICATION DATA:
30     APPLICATION NUMBER: US 08/286,889
31     FILING DATE: 5-AUG-1994
32     PRIOR APPLICATION DATA:
33     APPLICATION NUMBER: US 08/362,652
34     FILING DATE: 21-DEC-1994
35     ATTORNEY/AGENT INFORMATION:
36     NAME: Williams Jr., Joseph A.
37     REGISTRATION NUMBER: 38,659
38     REFERENCE/DOCKET NUMBER: 27866/32797
39     TELECOMMUNICATION INFORMATION:
40     TELEPHONE: 312-474-6300
41     TELEFAX: 312-474-0448
42     TELEX: 25-3856
43     INFORMATION FOR SEQ ID NO: 101:
44     SEQUENCE CHARACTERISTICS:
45     LENGTH: 413 amino acids
46     TYPE: amino acid
47     TOPOLOGY: linear
48     MOLECULE TYPE: protein
49     US-08-485-618-101

```

```

Query March      19.3%; Score 198; DB 1; Length 413;
Best Local Similarity 29.7%; Pred No 2, 5e-16;
Matches 60; Conservative 38; Mismatches 92; Indels 12; Gaps 8

OY      1 GICARVDASFOGSLAFTAQRCPY-MDVIVLDGNSI--YPMSEVQFTLRRLVGLKF 57
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      119 GFCVLDAHQPIGTGTPALPECPDQEMIVFLIDGSSISNDRPKKDKDFRAVMQD-F 177

OY      58 IDEBQIVGLVQGEPSFVHMSIGDPTREEVVRANKLSRREGREFTQAQIVNACTEG 117
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Db 177 KD-TNQFSLMGXNVNLVTHFTFSSFRSSNPDGLVEPIVQLTGL-TFTATGILKVVTEL 235

Qy 118 FSSHSGREARALILVWYTDGESH-DGSELPAALKACAGRVTFYGIIVAGHYLRQRDP 176

Db 236 FQYKNAKRSAKKILVITDQGYKDPRLHYSAVLPQAEQGITRYALGVDAF-----QKP 291

Qy 177 SFLREIRITIASDPDERFFENV 198

Db 292 TA-ROELDTIASPPAHVFOV 312

```

1      RESULT 8
2      US-08-605-672-101
3      ; Sequence 101, Application US/08605672
4      ; Patent No. 5817515
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Gallatin, W. Michael
7      ; APPLICANT: Van der Vieren, Monica
8      ; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
9      ; NUMBER OF SEQUENCES: 103
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
12     ; STREET: 233 South Wacker Drive, 6300 Seear Tower
13     ; City: Chicago
14     ; STATE: Illinois
15     ; COUNTRY: United States
16     ; ZIP: 60606-6402
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: Patentin Release #1.0, Version #1.25
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/08/605,672
24     ; FILING DATE:
25     ; CLASSIFICATION: 530
26     ; PRIOR APPLICATION DATA:
27     ; APPLICATION NUMBER: US 08/173,497
28     ; FILING DATE: 23-DEC-1993
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: US 08/286,889
31     ; FILING DATE: 5-AUG-1994
32     ; PRIOR APPLICATION DATA:
33     ; APPLICATION NUMBER: US 08/362,652
34     ; FILING DATE: 21-DEC-1994
35     ; ATTORNEY/AGENT INFORMATION:
36     ; NAME: Williams Jr., Joseph A.
37     ; REGISTRATION NUMBER: 38,659
38     ; REFERENCE/DOCKET NUMBER: 27866/32684
39     ; TELECOMMUNICATION INFORMATION:
40     ; TELEPHONE: 312-474-6300
41     ; TELEFAX: 312-474-0448
42     ; TELEX: 25-3856
43     ; INFORMATION FOR SEQ. ID NO.: 101:
44     ; SEQUENCE CHARACTERISTICS:
45     ; LENGTH: 413 amino acids
46     ; TYPE: amino acid
47     ; TOPOLOGY: linear
48     ; MOLECULE TYPE: protein
49     ; US-08-605-672-101

```

[illegible]

QY 118 FSQSHGGRPEARLLVVTDDGSH-DGEEPLPAALACAGRTVRYGIAVLGHYLRORDP 176  
Db 236 FQTKNGARESAAKLLIIVITDQCKYKDPHYSAVIPQAEQAGIRYAVIGVDAF----OKP 291  
QY 177 SSFLREIRTIASDPERFFENV 198  
Db 292 TA-ROELDTIASPEPDHAFVQV 312

## RESULT 9

US-08-482-293A-101  
; Sequence 101, Application US/08482293A  
; Patent No. 5831029  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5831029e1 Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,293A  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 101:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 413 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-482-293A-101

Query Match 19.3%; Score 198; DB 2; Length 413;  
Best Local Similarity 29.7%; Pred. No. 2.5e-16;  
Matches 60; Conservative 38; Mismatches 92; Indels 12; Gaps 8;

QY 1 GICARVDSFPOGSLAFTAORCPY--MDVIVILDGNSI--YPMSEVQFLRLVGLKF 57  
Db 119 GFCVILDHAAPIGIVPALPECPDQEMDIVFLIDSGSISNDIRKMKDFRAVMDQ-F 177  
QY 58 IDPEQIOVLQVGYGSPVHEWSLGDFTKEEVRAAKNLSRREGRETKTAQAINVACTEG 117  
Db 178 KD-TNTQSLMQYSNVLVHTFTFSSFRNSNPQGLVEPIVQLTGL-TFTATGILKVTEL 235  
QY 118 FSQSHGGRPEARLLVVTDDGSH-DGEEPLPAALACAGRTVRYGIAVLGHYLRORDP 176

Db 236 FQTKNGARESAAKLLIIVITDQCKYKDPHYSAVIPQAEQAGIRYAVIGVDAF----OKP 291  
QY 177 SSFLREIRTIASDPERFFENV 198  
Db 292 TA-ROELDTIASPEPDHAFVQV 312

## RESULT 10

US-08-943-363-101  
; Sequence 101, Application US/08943363  
; Patent No. 5837478  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5837478e1 Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,363  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 101:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 413 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-943-363-101

Query Match 19.3%; Score 198; DB 2; Length 413;  
Best Local Similarity 29.7%; Pred. No. 2.5e-16;  
Matches 60; Conservative 38; Mismatches 92; Indels 12; Gaps 8;

QY 1 GICARVDSFPOGSLAFTAORCPY--MDVIVILDGNSI--YPMSEVQFLRLVGLKF 57  
Db 119 GFCVILDHAAPIGIVPALPECPDQEMDIVFLIDSGSISNDIRKMKDFRAVMDQ-F 177  
QY 58 IDPEQIOVLQVGYGSPVHEWSLGDFTKEEVRAAKNLSRREGRETKTAQAINVACTEG 117  
Db 178 KD-TNTQSLMQYSNVLVHTFTFSSFRNSNPQGLVEPIVQLTGL-TFTATGILKVTEL 235  
QY 118 FSQSHGGRPEARLLVVTDDGSH-DGEEPLPAALACAGRTVRYGIAVLGHYLRORDP 176  
Db 236 FQTKNGARESAAKLLIIVITDQCKYKDPHYSAVIPQAEQAGIRYAVIGVDAF----OKP 291





STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 2072792  
US-08-897-443-3

Query Match 19.1%; Score 196; DB 2; Length 956;  
Best Local Similarity 30.2%; Pred. No. 1.7e-15;  
Matches 64; Conservative 33; Mismatches 83; Indels 32; Gaps 8;

QY 6 VDASQPOGSLAFLTQRCCTY-----MDVTVLDGNSI--YPMSEVQFTL 49  
DB 21 VGRERPOARFPRSGRHVMYDOTALLESSCENKADLVFIIDSSRSVNTYDAKKEFI 80  
QY 50 RLVLKLFIDPQIOVLQVGVESPVHEWSLGDFTKEVVARAKNLSRREGRETKTAQA 109  
DB 81 LDILQFLDIDPVTYVGLLOQSVKNEPSLKTFRKSEVERAVKRM-RHLSTGTMGLA 139  
QY 110 IMVACTEGFSQSHGRP---EAAFLLVVTDGSHDEELPALAKACEAGVTRYGIATL 166  
DB 140 IQYALNIAFSEAEAGAPLRENVPRIMITDGRPDQSAVEAA-KARNTG-ILIFALGV- 196  
QY 167 GHYLRQDPSSFLREIRTIADPDDEFFFNV 198  
DB 197 -----QVVD---LNTLKAIGSEPHKHDFLV 219

## RESULT 14

US-08-897-443-1  
Sequence 1, Application US/08897443  
Patent No. 5981263

## GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
APPLICANT: Kaser, Mathew  
TITLE OF INVENTION: HUMAN MATRILIN-3  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Inocyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA

ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/897,443  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0348 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 638 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: UTRSNOT02

CLONE: 681719  
US-08-897-443-1

Query Match 18.6%; Score 191; DB 2; Length 638;  
Best Local Similarity 33.0%; Pred. No. 3.8e-15;  
Matches 58; Conservative 33; Mismatches 67; Indels 18; Gaps 7;

QY 28 DVTIVLDGNSI--YPMSEVQFLRLVGLKFLIDPEQIQVLQVGVESPVHEWSLGDFT 85  
DB 57 DLVFIIDSSRSVNTYDAKKEFIYDILQFLDIDPVTYVGLLOQSVKNEPSLKTFR 116  
QY 86 KEVVRAAKNLSRREGRETKTAQAIMVACTEGFSQSHGRP---EAAFLLVVTDGSHD 142  
DB 117 KSEVERAVKRM-RHLSTGTMGLAIQYALNIAFSEAEAGAPLRENVPRIMITDGRPD 175  
QY 143 GEEPLPAKACEAGVTRYGIATVGHYLRQDPSSFLREIRTIADPDDEFFFNV 198  
DB 176 SVAEVA-KARDTG-ILIFALGV-----QVVD---FNTLSIGSEPHKHDFLV 219

## RESULT 15

US-08-199-776-2  
Sequence 2, Application US/08199776  
Patent No. 5594120

## GENERAL INFORMATION:

APPLICANT: Brenner, Michael B.  
APPLICANT: Parker, Christina M.  
TITLE OF INVENTION: No. 5594120e1 integrin alpha subunit  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield and Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: USA

ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/199,776  
FILING DATE:  
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1178 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-199-776-2

Query Match 17.3%; Score 177.5; DB 1; Length 1178;  
Best Local Similarity 31.0%; Pred. No. 5.1e-13;  
Matches 54; Conservative 31; Mismatches 80; Indels 9; Gaps 5;

QY 28 DVTIVLDGNSIYP--WSEVQFLRLVGLKFLIDPEQIQVLQVGVESPVHEWSLGDFT 85  
DB 201 EIALIILDSGSGSIDPDPFQARDFISNMNRNYEKCFCENFALVOYGVITQTEFLRDSOD 260  
QY 86 KEVVRAAKNLSRREGRETKTAQAIMVACTEGFSQSHGRPREARLVLVVTGDE-SHGE 144  
DB 261 VMASLARVQNTT-QVGSVTKTASAMQHVLDISFTSSHSRKRKAKVMVLTGGIPEDDL 319  
QY 145 ELPAALKACEAGVTRYGIATVGHYLRQDPSSFLREIRTIADPDDEFFFNV 198

Db 320 NLTVINSPKMOGVERFAIGV-GEFFKSART---ARELNLIASDPDETHAFKV 368

Search completed: July 16, 2003, 07:57:47  
Job time : 8.79747 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

```
Run on:      July 16, 2003, 08:41:16 ; Search time 40 Seconds
              (without alignments)
              475.865 Million cell updates/sec
```

Title: US-09-647-544-2\_COPY\_140\_337

Sequence: 1 GICARVDASFQPGSLAPTA.....FLREIRTIASDPDERFFENV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 100632

```
Minimum DB seq length: 0
Maximum DB seq length: 198
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :
1: pir_73:*
2: pir2:*
3: pir3:*
4: pir4:*
```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	336	32.8	191	2	147230	Vla-2 protein - p
2	74	7.2	194	2	151310	beta 8 integrin -
3	71	6.9	164	2	E75293	hypothetical prote
4	71	6.9	169	2	E71454	hypothetical prote
5	71	6.9	170	2	E161890	complement C4 - C
6	71	6.9	197	2	E850680	hypothetical prote
7	70.5	6.9	118	2	AF08660	hypothetical prote
8	70	6.8	188	2	AE33163	transcription regu
9	69.5	6.8	148	2	A98075	conserved hypot
10	69	6.7	129	2	C70666	hypothetical prote
11	68.5	6.7	148	2	G95210	acetyltransferase,
12	68	6.6	171	2	H75017	hypothetical prote
13	67.5	6.6	151	2	AF25820	hypothetical prote
14	67.5	6.6	193	2	E75544	jag-related protei
15	67	6.5	181	2	H81150	hypothetical prote
16	67	6.5	197	2	A69201	tuberythrin - Met
17	66.5	6.5	166	2	D83114	30S ribosomal pr
18	66.5	6.5	187	2	F86723	acetyl transferase
19	66	6.4	138	2	A82674	virulence regulat
20	66	6.4	180	2	D83656	conserved hypot
21	65.5	6.4	147	2	F71069	probable hydroge
22	65	6.3	86	1	KHIDS	gonadoliberin II
23	65	6.3	170	2	F137024	complement C4 - g
24	65	6.3	189	2	F98333	hypothetical prote
25	65	6.3	196	2	AD2949	acetyltransferase
26	64.5	6.3	190	2	H84412	hypothetical prote
27	64	6.2	129	2	S37713	hypothetical prote
28	64	6.2	129	2	AF3122	IS426 transposase
29	64	6.2	129	2	AG3158	IS 426 transposase

30	64	6.2	131	2	JC7739	paralytic peptide
31	62	6.2	157	2	T30772	hypothetical prot
32	63.5	6.2	85	2	A53453	gonadoliberin II
33	63.5	6.2	178	2	AH1330	hypothetical prot
34	63	6.1	194	2	H89940	conserved hypothe
35	62.5	6.1	108	2	B75029	hypothetical prot
36	62.5	6.1	178	2	AH1701	hypothetical prot
37	62	6.0	161	1	A30477	hemoglobin V1B-6
38	62	6.0	161	1	S04499	hemoglobin V1B-8
39	62	6.0	193	2	T32808	hypothetical prot
40	61.5	6.0	164	2	F84286	hypothetical prot
41	61.5	6.0	180	2	D83623	hypothetical prot
42	61.5	6.0	182	2	B56274	sulfur-regulated
43	61	6.0	102	2	D84397	hypothetical prot
44	61	6.0	161	1	UT0292	hemoglobin V1B-5
45	61	6.0	161	1	UT0349	hemoglobin V1B-5

## ALIGNMENTS

RESULT 1  
I47230

VLA-2 protein - pig (fragment)  
N:Alternate names: glycoprotein Ia/Ia  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 20-Sep-1999  
C:Accession: I47330; S21518  
R:Babou, W.F.; Potter, C.L.; Mirza, H.  
J Biol Chem 273:12741-12744, 1998

A:Title: The VIA-2 (alpha 2 beta 1) I domain functions as a ligand-specific recognition  
A:Reference number: 147230; PMID:7949129  
A:Accession: 147230  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: mRNA  
A:Residues: 1-191 <BAH>  
A:Cross-references: EMBL:Z12137; NID:G2158; PIDD:CAA78125.1; PIDD:G2159  
A:Superfamily: unassigned collagens  
A:Keywords: glycoprotein

Query Match 32.8%; Score 336; DB 2; Length 191;  
Best Local Similarity 43.9%; Pred. No. 1e-23;  
Matches 69; Conservative 27; Mismatches 61; Indels 0; Gaps 0

```

Qy 42 WSEVOFELRLVQKLFIDEOQOVGVYQYGBEPHMSJDPFRKKEEYVPAKXLSRREG 10
Db 1 WDAVKNLEKFFVOGJLDIGPTKQVLLIYANNPRVVFPLNLFKKAEMVEVTSHTTVOYG 60
Qy 102 REKTAQALMVAVCTBESGSGHGRCPEARLLVYVTDGESHGDELPALKKACEAGRTRY 16
Db 61 DLNTTKKALQYARDAVSAAAGRGCAIKVMVYVTDGESHGSMLKAVIDDQNDNITLRF 120
Qy 162 GIAVLGYLRQRDPSFLREIRTIASPDDEEFFNNV 198
Db 121 GIAVVLGYLRNALDITNLIKETKALASLPTERYFFNV 157

```

## RESULT 2

beta 8 integrin - chicken (fragment)  
C/Species: Gallus gallus (chicken)  
C/Date: 04-Sep-1997 #sequence\_revision 07-Nov-1997 #text\_change 20-Aug-1999  
C/Accession: I51310  
R/Venstrom, K.; Reichardt, L.  
Mol. Biol. Cell 6, 419-431, 1995  
A/Title: Beta 8 integrins mediate interactions of chick sensory neurons with laminin-1  
A/Reference number: I51310; MUID:95352850; PMID:7542940  
A/Accession: I51310  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-194 <VEN>  
A/Cross-references: NID:g1042047; PIDN:AAH34966.1; PID:g1042048  
C/Superfamily: integrin beta chain; laminin-type EGF-like homology



C:Keywords: 2Fe-2S; metalloprotein  
F:118,124,127,165/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 6.9%; Score 71; DB 2; Length 197;

Best Local Similarity 22.7%; Pred. No. 38;

Matches 27; Conservative 15; Mismatches 39; Indels 38; Gaps 5;

QY 94 KNLSSREG-----RETKQAQIMVA-----CTEGFSQSHGSGREPARLLVVTG 138

DB 29 KNLHRSYGHVQLSPVPRQARTSQEAWFLKSHKCTSTSSENGDETEKTIITLFPD- 87

QY 139 ESHDGEELPAALK-----ACEAGRV--TRYGIALVGHYLRORDPS 177

DB 88 --KDGEELIPVKVIGMSVLEAAHENDIDLEGACEASLACSTCHVIMDYMEVYNKLEEPT 144

#### RESULT 7

hypothetical protein stry3093 [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhimurium

A:Note: This species has also been called Salmonella typhimurium

C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001

C:Accession: AF0860

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, M.; Connor, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parky, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; PMID:11677608

A:Accession: AF0860

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-118 <PAR>

A:Cross-references: GB:AL513382; PDB:CAD06069.1; PID:g16504035; GSPDB:GN00176

C:Genetics:

A:Gene: stry3093

Query Match 6.9%; Score 70.5; DB 2; Length 118;

Best Local Similarity 21.4%; Pred. No. 23;

Matches 24; Conservative 15; Mismatches 24; Indels 49; Gaps 3;

QY 81 GDFRTKEVRAAKNLSREGRETKTAQIMVACTEGFSQSHGSGREPARLLVVTGDS 140

DB 47 GDYRQSEVIRSLRLREKQASR-IQLRLRLAEG--NSGEQA----- 90

QY 141 HDGEELPAALKACEAGRVTRYGIALVGHYLRORDPSFLREIRTIASDPDE 192

DB 91 -----WEKDAFLRVKTKGMIKDE 109

#### RESULT 8

transcription regulator, tetR family BMEI0891 [imported] - Brucella melitensis (strain 102/94)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: AE3363

R:DelVecchio, V.G.; Kaparakis, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova, E.; Mazur, M.; Goldtman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, M.; Gotsch, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentile, S.; Hamilton, N.; Holroyd, S.; Nardone, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whithead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: AV0500; PMID:9629597; PMID:9634230

A:Accession: C70686

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-129 <COL>

A:Cross-references: GB:Z81360; GB:AL123456; NID:g3261654; PDB:1A03704.1; PID:e276861;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV1720C

Query Match 6.7%; Score 69; DB 2; Length 129;

Best Local Similarity 23.0%; Pred. No. 35;

Matches 29; Conservative 19; Mismatches 50; Indels 28; Gaps 5;

Best Local Similarity 25.0%; Pred. No. 44;

Matches 28; Conservative 25; Mismatches 49; Indels 10; Gaps 5;

QY 85 TREEVRAAKNLSREGRETKTAQIMVACTEGFSQSHGSR-----PEARLLVVTGDS 140

DB 9 SREKTLRATLTAQGVGAHISLDL--VAARAGLSK--GGLLYSPPTAKLEAVKXVM 64

QY 141 HDGEELPAALKACEAGRVTRYGIALVGHY-LRORDPSFLREIRTIASDPD 191

DB 65 OEHEGAMVQETLQSGERNRVARAFLDYVRIOADKEPA--CGVLAALLENPD 115

#### RESULT 9

conserved hypothetical protein spr1627 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C:Accession: A98075

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; PMID:21429245; PMID:11544234

A:Accession: A98075

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-148 <KUR>

A:Cross-references: GB:AE007317; PDB:1AL00430.1; PID:g15459297; GSPDB:GN00174

C:Genetics:

A:Gene: spr1627

Query Match 6.8%; Score 69.5; DB 2; Length 148;

Best Local Similarity 25.7%; Pred. No. 37;

Matches 29; Conservative 23; Mismatches 34; Indels 27; Gaps 6;

QY 9 SFQPGSLAPTAQRCPYMDVIVLDGNSIYPWSEVOTFLRL-----VGKLFIDP 60

DB 27 SFQEQM-MGP---RIP-FLTLQLAEGVPSIFDQEPFGFKITLDEDSNLHIGFFNP 81

QY 61 EQLQVGL-----VOGESPVHWSLQDFRTKEVRAAKNLSREGRE 103

DB 82 QKQEGQLSGALRKRFVSLAFENEDISLVNF-----BANRAQLVYQKEGE 130

#### RESULT 10

hypothetical protein RV1720C - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: C70686

R:Coale, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;

Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentile, S.; Hamilton, N.; Holroyd, S.;

Nardone, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whithead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: AV0500; PMID:9629597; PMID:9634230

A:Accession: C70686

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-129 <COL>

A:Cross-references: GB:Z81360; GB:AL123456; NID:g3261654; PDB:1A03704.1; PID:e276861;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV1720C

Query Match 6.7%; Score 69; DB 2; Length 129;

Best Local Similarity 23.0%; Pred. No. 35;

Matches 29; Conservative 19; Mismatches 50; Indels 28; Gaps 5;

QY 12 PQSLAPTAQR-----CPTYMDVIVLDGNSIYPWSEVOTFLRLVGLKLFIDPEIQ 64

```

Db      17 PAC--AAVARRLGETVHAFAHFDV-----EVIGAIRQAAVRQQLISDHEGL 60
QY      65 VGLVQGESFVHWSLGDFTKEEVVRAAKNLSRREGRETKTAQAL---MVACTEGFSQS 121
Db      61 VVVVNFLLSLFVRRLPKPFTQRAYQURSTHTVA--DGAVVALAELGLVPLITCDGLAOS 118
QY      122 HGGRIPE 127
Db      119 HGHNAE 124

RESULT 11
G95210
acetyltransferase, GNAT family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: G95210
R:Retellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95210
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <KUR>
A:Cross-references: GB:AE005672, PIDN:AAK75880.1, PID:g14973106, GSPDB:GN00164, TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1807

Query Match
Best Local Similarity 25.0%; Score 68.5; DB 2; Length 148;
Pred. No. 45;
Matches 28; Conservative 25; Mismatches 34; Indels 25; Gaps 6;

QY      9 SFPOGSLAPTAQRCPYMDVIVLDGNSIYPMSEVQTFRLRL-----VGKLFIDP 60
Db      27 SFQEQW-MGP---RIP-FLITLQALGCVSIFDEQFVGFIQIKRLSDSNLHGRFIP 81

QY      61 EQLQVGLVQGESFVHWSLGDFTKE-----EVVRAAKNLSRREGRE 103
Db      82 OKQEQQL--GSKALRKRFVSLAENRDISLSLNVAEANORQNLVQKGGFE 130

RESULT 12
H75017
hypothetical protein PAB1243 - Pyrococcus abyssi (strain Oxa5)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H75017
R:anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: H75017
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50582.1; PID:e151648
A:Experimental source: strain Oxa5
C:Genetics:
A:Gene: PAB1243

Query Match
Best Local Similarity 24.3%; Score 68; DB 2; Length 171;
Pred. No. 60;
Matches 33; Conservative 17; Mismatches 44; Indels 42; Gaps 6;

QY      40 YKMSVQOTLRLVGLTFLDPEQIQVGLVQGESF-----VHWSLGDGPD--- 83
Db      52 YKISHISSAMKYLEG-----VGLVQRIKKPGDRRAVATKPNFSEWSSAFYEKI 101

```

```

QY      84 -----RTKEEVVRAAKNLSRREGRETKTAQALMVACTEGFSQSHGRIPEAR-LIIVVTD 137
Db      102 LRDIQETESIMRALAELEEGEGSEVEIEIKETLKAALR-----RNEVRAKLLTLIMQ 153
QY      138 GESHDGEELPAALKAC 153
Db      154 FKSE--EELIKVLESC 167

RESULT 13
AF2520
hypothetical protein alr7342 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AF2520
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasa moto, S.; Matanabe, A.; Iritguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2520
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA077100.1; PID:g17134541; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr7342
A:Genome: plasmid

Query Match
Best Local Similarity 26.6%; Score 67.5; DB 2; Length 151;
Pred. No. 58;
Matches 14; Conservative 19; Mismatches 18; Indels 1; Gaps 1;

QY      50 RRLVGLKFLIDPEQIQVGLVQGESFVH--EWSLGDFTKEEVVRAAKNLSRRE 100
Db      95 KRLVAVEVFNRRSVLTVVQGGAAVVRYKLGKYNTRKEQFLQEAANKQK 146

RESULT 14
E75544
Jag-related protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: E75544
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.D.;
M.; Shen, M.; Vamathevan, J.D.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75544
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-193 <WHI>
A:Cross-references: GB:AF001866; GB:AE000513; NID:g6457921; PIDN:AAF09828.1; PID:g645791
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0246
A:Map position: 1

Query Match
Best Local Similarity 24.5%; Score 67.5; DB 2; Length 193;
Pred. No. 77;
Matches 40; Conservative 22; Mismatches 50; Indels 51; Gaps 9;

QY      13 QGSLAPTAQRCPYMDVIVLDGNSIYPMSEVQTFRLRLVGLKFLIDPE-QIQVGLVQYG 71
Db      21 ESALPPRA---PDAPVSAPIHGADG-DRAVLQGLFLAELAR--IDPGLRVQV----- 68
QY      72 ESPVHWSLGDFTKEEVVRA--AKNLSRREGRETKTAQALMVACTEGFSQSHGRIPE- 127

```

DB 69 -----RETEDALEAEISGENNAARLAGRDCRTGAEIVAYAVLAK-HAGRGDL 115  
QY 128 -----AARLLV-VYTDGESHDEELPAA 149  
DB 116 RYRVDVCGFRKROADTLTKLARLAVQVAKSGEPHELQPMPPA 158

## RESULT 15

H81150  
hypothetical protein NMB0861 [imported] - Neisseria meningitidis (strain MC58 serogroup  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Feb-2001  
C:Accession: H81150; F81872  
R:Letellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, S.D.; Dougherty, B.A.;  
rt, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: H81150  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-181 <TET>  
A:Cross-References: GB:AEO02439; GB:AEO02098; NID:g7226100; PIDN:AAFA1272.1; PID:g722609  
A:Experimental source: serogroup B, strain MC58  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:2022556; PMID:10761919  
A:Accession: F81872  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-181 <PAR>  
A:Cross-References: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84336.1; PID:g737976  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMB0861; NMA1073

Query Match 6.5%; Score 67; DB 2; Length 181;  
Best Local Similarity 23.5%; Pred.No. 80;  
Matches 28; Conservative 17; Mismatches 48; Indels 26; Gaps 3;  
QY 83 FRTKEVVPAKNTLSRREGRETKTA-----QAIWVACTEGFSGSHGRPEARLTV 133  
DB 33 FHTTRADAPWQALAEISGKEMKTEGAFLPLAIIIGAAIGMTQHGFSYATTGKRPASVRYVA 92  
QY 134 V-----VTGESHDEELPAA-----LRACEAGRVTRYGIAVLGHYLRORD 175  
DB 93 IAGGIGAIKPGVGAAGKVVSPFAKYGREIKIGNNMRIAPFGNRTGHPICKKFPYHRRVTD 151

Search completed: July 16, 2003, 08:45:47  
Job time : 42 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:34:45 ; Search time 23 Seconds  
(without alignments)  
357.057 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025  
Sequence: 1 GICARVDAFQPOGSLAPTA.....FLREIRITASDPDERFFENV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 36360

Minimum DB seq length: 0  
Maximum DB seq length: 198

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	6.6	196	1 YEE9_STRCO	O9FBN0 streptomyce
2	65	6.3	86	1 GON2_CLAGA	P43306 clarias gar
3	63.5	6.2	85	1 GON2_HAPBU	P37044 haplochromi
4	63.5	6.2	150	1 HIS3_RHIME	O928B5 rhizobium m
5	62	6.0	161	1 GLBC_CHITH	P12548 chironomus
6	62	6.0	161	1 GLBF_CHITH	P13549 chironomus
7	62	6.0	161	1 GLBI_CHITH	O23763 chironomus
8	62	6.0	161	1 GLBI_CHITP	P18966 chironomus
9	62	6.0	161	1 GLBK_CHITH	O23762 chironomus
10	61.5	6.0	182	1 SRPB_SYN7	O55026 synchococc
11	61	6.0	161	1 GLBE_CHITH	P15833 chironomus
12	60.5	5.9	85	1 GON2_DICLA	O40158 lycopersico
13	60.5	5.9	85	1 GON2_DICLA	O91808 dicentrarch
14	60.5	5.9	96	1 UTER_RAT	P17559 ratius novy
15	60.5	5.9	140	1 V250_FOWPV	P14562 fowlpox vir
16	60	5.9	129	1 Y207_ABRPE	O9YF22 aecopyrum p
17	60	5.9	181	1 RL10_ANASP	O8YJ16 anabaena sp
18	60	5.9	194	1 YCEP_SALTY	P58627 salmoneilla
19	60	5.9	198	1 RNM2_SALTY	P40675 salmoneilla
20	59.5	5.8	91	1 DBH_LACLA	O96164 lactococcus
21	59.5	5.8	140	1 NDK_METKA	O98661 methanococc
22	59.5	5.8	159	1 Y399_METKA	P58829 methanopyru
23	59.5	5.8	169	1 ILVH_MYCLE	O33113 mycobacteri
24	58.5	5.7	86	1 GON2_CARAU	P51924 carassius a
25	58.5	5.7	142	1 KLA1_ECOLI	P52602 escherichia
26	58.5	5.7	145	1 GLB7_CHITH	P02226 chironomus
27	58.5	5.7	186	1 RELX_RAT	P01347 ratius novy
28	58	5.7	95	1 CH10_BORPE	P48821 bordetella
29	58	5.7	131	1 YM24_PYPAB	O9V280 pyrococcus
30	58	5.7	141	1 NDK_ABRPE	O9V280 aecopyrum p
31	58	5.7	167	1 ILVH_MYCAV	O59499 mycobacteri
32	58	5.7	179	1 RML6_AATH	O95747 atabadiopsis
33	58	5.7	180	1 UCRI_SYN2	P26292 synchococc

34	57.5	5.6	85	1 GON2_MORSA	O73811 morone saxa
35	57.5	5.6	150	1 HIS3_AGRIS	O8uek7 agrobacteri
36	57.5	5.6	170	1 DAPB_COXBU	P24703 coxiella bu
37	57.5	5.6	174	1 RUVG_PSEAE	O51424 pseudomonas
38	57.5	5.6	177	1 RUG_VIRCH	O9knz9 vibrio chol
39	57.5	5.6	182	1 PYRE_STRCO	O9x817 streptomyce
40	57.5	5.6	193	1 LEPT_BACAM	P41025 bacillus am
41	57.5	5.6	195	1 DCD_HALNI	O9hsg3 halobacteri
42	57	5.6	101	1 IL8_RABIT	P19874 oryctolagus
43	57	5.6	112	1 YWFM_ECOLI	P75977 escherichia
44	57	5.6	113	1 MRA4_ENTHI	P16877 entameba h
45	57	5.6	164	1 RL17_TREPA	O83243 treponema p

## ALIGNMENTS

```

RESULT 1
ID YEE9_STRCO STANDARD; PRT; 196 AA.
AC O9FBN0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative NADH dehydrogenase/NAD(P)H nitroreductase SC05049 (EC
DE 1.-.-.-).
GN SC05049 OR SCX7.22C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.M.A.
RC STRAIN=AJ3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wetzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor AJ3(2)";
RL Nature 417:141-147(2002).
CC -I- SIMILARITY: BELONGS TO THE NITROREDUCTASE FAMILY. HADB SUBFAMILY.
CC -I- SIMILARITY: BELONGS TO THE NITROREDUCTASE FAMILY. HADB SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AL391754; CAC05894.1;
DR InterPro, IPR00415; Nitroreductase.
DR Pfam, PF00881; Nitroreductase; 1.
KW Hypothetical protein; Oxidoreductase; NAD; NADP; Flavoprotein; FMN;
KW Complete proteome.
SQ
SEQUENCE 196 AA; 21473 MW; F29AF0C8830F954B CMC64;
Query Match 6.6%; Score 68; DB 1; Length 196;
Best Local Similarity 25.2%; Pred. No. 34;
Matches 37; Conservative 14; Mismatches 70; Indels 26; Gaps 4;
OY 31 ILVGSNSIYPMSEVQFLRLVGLKFLPD-----PEQIQV--GIYOGSPPHMSIGDGR 84
DB 3 LVLD-----PAADQLLREARANTNTFDEPVTDEOVQAIYDLVKGPTAFNQSPLRITL 56

```

QY 85 TKEEVRAAKNLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVTGDSHDE 144  
 Db 57 VNSPEARRELVAMAGNRPKTAAPLVALLSADNFBELPBLF-----P 102  
 QY 145 ELPAALACEAGRVTRYGIAVLGHYLR 171  
 Db 103 HFPAAKDAPFESRPVREGAATLNAALQ 129

## RESULT 2

GON2\_CLAGA STANDARD; PRT; 86 AA.  
 ID GON2\_CLAGA  
 AC P43306;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Gonadoliberin II precursor (Gonadotropin-releasing hormone II)  
 DE (GNRH-II) (LH-RH II) (Lutiberin II).  
 GN GNRH2.  
 OS Clarias gariepinus (Sharptooth catfish) (African catfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;  
 OC Clariidae; Clarias.  
 OC NCBI\_TaxID=13013;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94291651; PubMed=8020492;  
 RA Bogerd J., Zandbergen T., Andersson E., Goos H.,  
 RT "Isolation, characterization and expression of cDNAs encoding the  
 catfish-type and chicken-II-type gonadotropin-releasing-hormone  
 RT precursors in the African catfish".  
 RL Eur. J. Biochem. 222:541-549(1994).  
 RN [2]  
 RP SEQUENCE OF 25-34.  
 RC TISSUE=Brain;  
 RX MEDLINE=92392313; PubMed=1520292;  
 RA Bogerd J., Li K.W., Janssen-Dommerholt C., Goos H.,  
 RT "Two gonadotropin-releasing hormones from African catfish (Clarias  
 gariepinus)".  
 RL Biochem. Biophys. Res. Commun. 187:127-134(1992).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: X78047; CAA54969.1; -  
 DR PIR: JCI243; RHID2S.  
 DR InterPro: IPR002012; GNRH.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KM Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KW Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 86 PROGNADOLIBERIN II.  
 FT PEPTIDE 25 34 GONADOLIBERIN II.  
 FT PEPTIDE 38 86 GNRH-ASSOCIATED PEPTIDE II.  
 FT MOD\_RES 25 25 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 34 34 AMIDATION (G-35 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 86 AA; 9766 MW; 4AD9F24597E77EBF CRC64;

Query Match 6.3%; Score 65; DB 1; Length 86;  
 Best Local Similarity 26.8%; Pred. No. 24;  
 Matches 22; Conservative 10; Mismatches 20; Indels 30; Gaps 4;  
 QY 109 AIMVACTE---GFSQ--SHGGRPEARLLVVTGDSHDEELPALAKACEAGRVTRYGI 163

Db 11 AALLLCQAQLSFSGHWSHGWPGGKREI-----DSVSSPEISGEIKLCAGE----- 58  
 QY 164 AVLGHYLRORDPSSFLREIR 185  
 Db 59 -----CSYLRPLRT 67

## RESULT 3

GON2\_HAPBU STANDARD; PRT; 85 AA.  
 ID GON2\_HAPBU  
 AC P37044; P20408;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Gonadoliberin II precursor (Gonadotropin-releasing hormone II)  
 DE (GNRH-II) (LH-RH II) (Lutiberin II).  
 GN GNRH2.  
 OS Haplochromis burtoni (Burton's mouthbrooder).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorphi; Acanthopterygii; Percomorpha; Perciformes; Labroidae;  
 OC Cichlidae; Ascotilapia.  
 OC NCBI\_TaxID=8153;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94151343; PubMed=8108425;  
 RA White S.A., Bond C.T., Francis R.C., Kaeten T.L., Fernald R.D.,  
 RA Adelman J.P.;  
 RT "A second gene for gonadotropin-releasing hormone: cDNA and  
 RT expression pattern in the brain".  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1423-1427(1994).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ONLY ONE CELL GROUP IN THE  
 CC MESENCEPHALON.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: L27435; AAA74993.1; -  
 DR InterPro: IPR002012; GNRH.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KM Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KW Signal; Multigene family.  
 FT SIGNAL 1 23  
 FT CHAIN 24 85 PROGNADOLIBERIN II.  
 FT PEPTIDE 24 33 GONADOLIBERIN II.  
 FT PEPTIDE 37 64 GNRH-ASSOCIATED PEPTIDE II-1 (POTENTIAL).  
 FT PEPTIDE 67 85 GNRH-ASSOCIATED PEPTIDE II-2 (POTENTIAL).  
 FT MOD\_RES 24 24 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 85 AA; 9631 MW; CF8C0EBDF27365F CRC64;

Query Match 6.2%; Score 63.5; DB 1; Length 85;  
 Best Local Similarity 32.9%; Pred. No. 32;  
 Matches 25; Conservative 6; Mismatches 26; Indels 19; Gaps 4;  
 QY 110 IMVACTEFSQ--SHGGRPEARLLVVTGDSHDEELPALAKACEAGRVTRYGIAVLG 167  
 Db 14 LCVGAQGSFAQHMWSHGWPGGKREL-----DSFGTSSISEIKLCAGEGS----- 59  
 QY 168 HYLRORDPSSFLREI 183  
 Db 60 -YLRPQ--RSILRNI 72



OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
 OC Chironomidae; Chironomidae; Chironominae; Chironomus.  
 OK NCBI\_TaxID=7155;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89137998; PubMed=2852146;  
 RA Trewitt P.M., Saffarini D.A., Bergtrom G.;  
 RT "Multiple clustered genes of the haemoglobin VIIB subfamily of  
 RT Chironomus thummi thummi (Diptera).";  
 RL Gene 69:91-100(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96054034; PubMed=7563117;  
 RA Trewitt P.M., Luhm R.A., Samad F., Ramakrishnan S., Kao W.Y.,  
 RA Bergtrom G.;  
 RT "Molecular evolutionary analysis of the YWVZ/7B globin gene cluster  
 RT of the insect Chironomus thummi.";  
 RL J. Mol. Evol. 41:313-328(1995).  
 RN [3]  
 RP SEQUENCE OF 17-161 (MIXTURE OF ISOZYMES).  
 RX MEDLINE=79129112; PubMed=422121;  
 RA Sladic-Simic D., Kleinschmidt T., Braunitzer G.;  
 RT "Hemoglobins, XXVI. Analysis of the primary structure of the dimeric  
 RT insect haemoglobin CTT VIIb (Erythrocytorin) from Chironomus thummi  
 RT thummi, Diptera.";  
 RL Hoppe-Seiler's Z. Physiol. Chem. 360:115-124(1979).  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST 12 DIFFERENT COMPONENTS IN MIDGE  
 CC GLOBIN.  
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST NINE GENES FOR VIIB VARIANTS.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U07703; AAA85486.1; -  
 DR PIR: A02547; G01CE8.  
 DR PIR: A30477; A30477.  
 DR HSSP: P02229; 1ECA.  
 DR InterPro: IPR002336; Erythruin.  
 DR InterPro: IPR000971; Globin.  
 DR Pfam: PF00042; globin.1.  
 DR PRINTS: PR00611; ERYTHROURIN.  
 DR PROSITE: PS01033; GLOBIN.1.  
 KW Heme; Oxygen transport; Transport; Multigene family;  
 KM Signal.  
 FT SIGNAL. 1 16  
 FT CHAIN 17 161 GLOBIN CTT-VIIB-6.  
 FT METAL 76 76 IRON (HEME DISTAL LIGAND)  
 FT (BY SIMILARITY).  
 FT METAL 111 111 IRON (HEME PROXIMAL LIGAND)  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 161 AA; 16959 MW; EA1B8C54BA432EF CRC64;  
 Query Match 6.0%; Score 62; DB 1; Length 161;  
 Best Local Similarity 29.5%; Pred. No. 95;  
 Matches 23; Conservative 10; Mismatches 29; Indels 16; Gaps 2;  
 Oy 8 ASFOGGSIAPTAQCPTMDVIVLDGNSIYPMSEVOTFLRLVYGLFIDPEIOVGL 67  
 Db 65 ASIKOTGAFATHTATVTSFLSEVIALSGNDS-----NAAVNSLVSKLGDHKGAVSA 118  
 Oy 68 VOYGESPVHEWSLGDERT 85  
 Db 119 AQFG-----FRT 126

RESULT 7  
 GLOBI CHITP STANDARD; PRT; 161 AA.  
 AC 023763;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Globin CTT-VIIB-8 precursor.  
 GN CTT-7B8.  
 OS Chironomus thummi thummi (Midge).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
 OC Chironomidae; Chironomidae; Chironominae; Chironomus.  
 OK NCBI\_TaxID=7155;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96054034; PubMed=7563117;  
 RA Trewitt P.M., Luhm R.A., Samad F., Ramakrishnan S., Kao W.Y.,  
 RA Bergtrom G.;  
 RT "Molecular evolutionary analysis of the YWVZ/7B globin gene cluster  
 RT of the insect Chironomus thummi.";  
 RL J. Mol. Evol. 41:313-328(1995).  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST 12 DIFFERENT COMPONENTS IN MIDGE  
 CC GLOBIN.  
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST NINE GENES FOR VIIB VARIANTS.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U07703; AAA85490.1; -  
 DR HSSP: P02229; 1ECA.  
 DR InterPro: IPR002336; Erythruin.  
 DR InterPro: IPR000971; Globin.  
 DR Pfam: PF00042; globin.1.  
 DR PRINTS: PR00611; ERYTHROURIN.  
 DR PROSITE: PS01033; GLOBIN.1.  
 KW Heme; Oxygen transport; Transport; Multigene family;  
 KM Signal.  
 FT SIGNAL. 1 16  
 FT CHAIN 17 161 GLOBIN CTT-VIIB-8.  
 FT METAL 76 76 IRON (HEME DISTAL LIGAND)  
 FT (BY SIMILARITY).  
 FT METAL 111 111 IRON (HEME PROXIMAL LIGAND)  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 161 AA; 17040 MW; C7BB9E70C1F7E161 CRC64;  
 Query Match 6.0%; Score 62; DB 1; Length 161;  
 Best Local Similarity 29.5%; Pred. No. 95;  
 Matches 23; Conservative 10; Mismatches 29; Indels 16; Gaps 2;  
 Oy 8 ASFOGGSIAPTAQCPTMDVIVLDGNSIYPMSEVOTFLRLVYGLFIDPEIOVGL 67  
 Db 65 ASIKOTGAFATHTATVTSFLSEVIALSGNDS-----NAAVNSLVSKLGDHKGAVSA 118  
 Oy 68 VOYGESPVHEWSLGDERT 85  
 Db 119 AQFG-----FRT 126  
 RESULT 8  
 GLOBI CHITP STANDARD; PRT; 161 AA.  
 AC P18966;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

DE Globin CTT-VIIB-8 precursor.
OS Chironomus thummi piger (Midge).
OC Eukaryota, Metazoa; Arthropoda, Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7156;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89374601; PubMed=2775460;
RA Rozynek P., Hankeln T., Schmidt E.R.;
RT "Structure of a hemoglobin gene cluster and nucleotide sequence of
RT three hemoglobin genes from the midge Chironomus thummi piger
RT (Diptera, Insecta).";
RL Biol. Chem. Hoppe-Seyler 370:523-542(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Hankeln T., Rozynek P., Schmidt E.R., Broecker M.;
RL Submitted (SEP-1990) to the EMBL/GenBank/DBD databases.
CC -1- SUBUNIT: Homodimer.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST 12 DIFFERENT COMPONENTS IN MIDGE
CC GLOBIN.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST NINE GENES FOR VIIB VARIANTS.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X56271; CAA39713.1; -.
DR PIR, S04499; S04499.
DR PIR, S21628; S21628.
DR HSSP, P02229; IECA.
DR InterPro, IPR002336; Erythcrutin.
DR InterPro, IPR000971; Globin.
DR Pfam, PF00042; globin.1.
DR PRINTS; PR00611; ERYTHCRURIN.
DR PROSITE; PS01033; GLOBIN.1.
DR KW Heme; Oxygen transport; Transport; Multigene family;
FT SIGNAL.
FT SIGNAL 1 16 GLOBIN CTT-VIIB-8.
FT CHAIN 17 161 IRON (HEME DISTAL LIGAND)
FT METAL 76 76 (BY SIMILARITY)
FT FT METAL 111 111 IRON (HEME PROXIMAL LIGAND)
FT FT (BY SIMILARITY).
SQ SEQUENCE 161 AA; 16999 MW; 048B981F5528C373 CRC64;

Query Match 6.0%; Score 62; DB 1; Length 161;
Best Local Similarity 29.5%; Pred. No. 95;
Matches 23; Conservative 10; Mismatches 29; Indels 16; Gaps 2

OY 8 ASFCQGSGLAPACRCPTVMVVILYDGSNSIYPMSEVOTFLRLVGLKFLIPEDIQVGL 67
DB 65 ASIKGTGAFATATATIVFSLVSLVDSGNS-----NAAVNSLVSKLDGDKRGVSA 118
OY 68 VOYGESPVHEWSLGDFT 85
DB 119 AQFGE-----FRT 126

RESULT 9
ID GLOB_CHTH STANDARD; PRT; 161 AA.
AC Q23762;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Globin CTT-VIIB-10 precursor.

```

```

GN CTT-7B10
OS Chironomus thummi thummi (Midge).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7155;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96054034; PubMed=7563117;
RA Trewitt P.M., Lohm R.A., Samad F., Ramakrishnan S., Kao W.Y.,
RA Bergstrom G.;
RT "Molecular evolutionary analysis of the ywvz/7b globin gene cluster
RT of the insect Chironomus thummi.";
RL J. Mol. Evol. 41:313-328 (1995).
CC -1- SUBUNIT: Homodimer.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST 12 DIFFERENT COMPONENTS IN MIDGE
CC GLOBIN.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST NINE GENES FOR VIIB VARIANTS.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; U07703; AA85488.1; -.
DR HSP; P02329; IBCA.
DR InterPro; IPR002336; Erythecruin.
DR InterPro; IPR000971; Globin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PRO0611; ERYTHCRUORIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Multigene family;
KW Signal.
FT SIGNAL. 1 16
FT CHAIN 17 161 GLOBIN CTT-VIIB-10.
FT METAL 76 76 IRON (HEME DIGITAL LIGAND)
FT METAL 111 111 (BY SIMILARITY).
FT METAL 111 111 IRON (HEME PROXIMAL LIGAND)
FT FT (BY SIMILARITY).
SQ SEQUENCE 161 AA; 16989 MW; 077BA4EC9A14334C CRC64;
Query Match 6.0%; Score 62; DB 1; Length 161;
Best Local Similarity 29.5%; Pred. No. 95;
Matches 23; Conservative 10; Mismatches 29; Indels 16; Gaps 2
QY 8 ASFQPGSLAFTACRPTVMYDVIVLDGNSIYPMSEVQTFRLRLVGLFTDPEQIOVGL 67
DB 65 ASIKDTGAFATNARIVSFLSEVIALSGNES-----NAAVNSLVSKLGDHKARAGVA 118
QY 68 VQGESPVHEWSLGDFT 85
DB 119 AQFGB-----FRF 126
RESULT 10
SRFB_SYNP7
ID SRFB_SYNP7 STANDARD; PRT; 182 AA.
AC Q55026;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SrpB protein.
GN SrpB.
OS Synechococcus sp. (strain PCC 7942) (Anacyetis nidulans R2).
OC Plasmid pNLT.
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.

```



```

RESULT 12
MT2B_LYCSES STANDARD; PRT; 82 AA.
ID MT2B_LYCSES
AC 040158; 043514;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Metallothionein-like protein type 2 B.
GN MTB.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxId=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ev. Ailea Craig;
RA Whitehaw C.A.; Lénuquet J.A.; Thurman D.A.; Tomsett A.B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bonner Beste; TISSUE=Root;
RA Girtch A., Herdik A., Balzer H., Stephan U., Baumlein H.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEIN HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U7966; AAB04675.1; -
DR EMBL; Z68138; CAA92243.1; -
DR InterPro: IPR000347; Metallothion_15.
DR Pfam: PF01439; Metallothio_2; 1.
DR ProDom: PD001611; Metallothion_15; 1.
DR Metal-binding; Metal-thiolate cluster; Multigene family.
FT CONFLICT 17 17 G -> D (IN REF. 2).
SQ SEQUENCE 82 AA; 8253 MW; 6B298C4915CE495F CRC64;

Query Match 5.9%; Score 60.5; DB 1; Length 82;
Best Local Similarity 37.0%; Pred. No. 57;
Matches 17; Conservative 7; Mismatches 17; Indels 5; Gaps 2;

QY 35 GSNSTYIP---MSGVQFLRLVGLFIDPEQIVGVGVGSGSPVHE 77
Db 21 GGCNYPDMSTYESTTETTV--LGVGPEKTSFGAMGESPVAE 64

RESULT 13
GN2_DICLA STANDARD; PRT; 85 AA.
ID GN2_DICLA
AC Q91A08;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gonadotropin II precursor (Gonadotropin-releasing hormone II)
DE (GNRH-II) (LH-RH II) (Laliberin II).
GN GNRH2.
OS Dicotylarchus labrax (European sea bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Percoidae;
OC Moronidae; Dicentrarchus.
OX NCBI_TaxId=13489;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Brain;
RX MEDLINE=20540016; PubMed=11086295;
RA Gonzalez-Martinez D., Madigou T., Zmora N., Anglade I., Zanny S.,
RA Zohar Y., Elizur A., Munoz-Cuelo J.A., Kah O.;
RT "Differential expression of three different prepro-GNRH
RT (gonadotropin-releasing hormone) messengers in the brain of the
RT European sea bass (Dicentrarchus labrax).";
RL J. Comp. Neurol. 429:144-155(2001).
CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY
CC SIMILARITY).
CC -----
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF224281; AAF62900.1; -
DR InterPro: IPR002012; GNRH.
DR Pfam: PF00446; GNRH; 1.
DR PROSITE: PS00473; GNRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Signal.
FT SIGNAL 1 23
FT CHAIN 24 85 BY SIMILARITY.
FT PEPTIDE 24 33 GONADOLIBERIN II.
FT PEPTIDE 37 85 GNRH-ASSOCIATED PEPTIDE II.
FT MOD_RES 24 24 PYROLIDONE CARBOXYLIC ACID
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP) (BY
FT MOD_RES 33 33 SIMILARITY).
SQ SEQUENCE 85 AA; 9646 MW; F832C0698C942C64 CRC64;

Query Match 5.9%; Score 60.5; DB 1; Length 85;
Best Local Similarity 34.9%; Pred. No. 59;
Matches 22; Conservative 4; Mismatches 20; Indels 17; Gaps 3;

QY 121 SHGSPERARLLVVTGDSHGDELPAALACAGRVTRYGLVGLHYLRQDPSSFL 180
Db 27 SHGWPGGKREL-----DSFGTSEISBEIKTCGEGCS-----YLPQR--RSVL 69

QY 181 REI 183
Db 70 RNI 72

RESULT 14
UTER_RAT STANDARD; PRT; 96 AA.
ID UTER_RAT
AC P17559;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Clara cell phospholipid-binding protein precursor (CCBPp) (Clara cells
DE 10 kDa secretory protein) (CC10) (Uteroglobin) (PCB-binding protein).
GN SCGB1A1 OR UGB OR CC10 OR UTG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Katyal S.L., Singh G., Brown W.E., Kennedy A.L., Squeglia N.,
RA Wong-Chong M.-L.;
RT "Clara cell secretory (10 kDa) protein: amino acid and cDNA
RT nucleotide sequences and developmental expression.";
RL Prog. Respir. Res. 25:29-35(1990).
RN [2]
RP SEQUENCE FROM N.A.

```

[illegible][illegible]

RESULT 15	2V50_FOWPV	STANDARD;	PRT; 140 AA.
ID	_V250_FOWPV		
AC	P14362;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical protein FPV250 (BamHI-ORF4).		
GN	FPV250.		
OS	Fowlpox virus (FPV).		
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;		
OC	Avipoxvirus.		
OX	NCBI_TaxId=10261;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FP-9 / Isolate HP-438;		
RX	MEDLINE=86229622; PubMed=2836548;		
RA	Tomley F., Binn M., Campbell J., Boursnell M.E.G.;		
RT	"Sequence analysis of an 11.2 kilobase, near-terminal, BamHI fragment		
RT	of fowlpox virus.";		
RL	J. Gen. Virol. 69:1025-1040(1988).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20193820; PubMed=10729155;		
RA	Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;		
RT	"The genome of fowlpox virus.";		
RL	J. Virol. 74:3815-3831(2000).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its use by		
CC	non-profit institutions as long as its content is in no way used for commercial		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; D00295; BAA00196.1; -;		
DR	EMBL; AF198100; AAF44594.1; -;		
DR	PIR; D29963; MWVZF4.		
KW	Hypothetical protein; Early protein.		
SO	SEQUENCE 140 AA; 16507 MW; 09094006BC510C425 CRC64;		

[illegible]



Wed Jul 16 08:48:45 2003

us-09-647-544-2\_copy\_140\_337.closed.rsp

Page 9

Job time : 25 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:40:06 ; Search time 79 Seconds

(without alignments)  
516.422 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025  
Sequence: 1 GICARVDASFQPGSLAPTA.....FLREIRIASDPDERFFENV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 284930

Minimum DB seq length: 0

Maximum DB seq length: 198

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mmc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_proteint.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriaph.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	336	32.8	191	6	Q29124	Q29124 sue scrofa
2	92	9.0	146	5	O96930	O96930 plasmidium
3	74	7.2	194	13	Q91415	Q91415 gallus gall
4	72.5	7.1	134	11	Q920X7	Q920X7 cavia porce
5	71.5	7.0	108	17	O8TW99	O8TW99 methanopyru
6	71	6.9	164	16	O9RS46	O9RS46 deinococcus
7	71	6.9	169	17	O58027	O58027 pyrococcus
8	71	6.9	170	6	Q28796	Q28796 pan troglod
9	71	6.9	174	16	Q9X957	Q9X957 streptomyce
10	70.5	6.9	197	10	O9M0V0	O9M0V0 arabidopsis
11	70.5	6.9	118	16	O8ZME3	O8ZME3 salmonella
12	70.5	6.8	118	16	O8Z448	O8Z448 salmonella
13	70	6.8	188	16	O8YH81	O8YH81 brucella me
14	69	6.7	129	16	P71978	P71978 mycobacteri
15	68.5	6.7	148	16	O97P37	O97P37 streptococc
16	68.5	6.7	163	4	O9HB01	O9HB01 homo sapien

17	68	6.6	133	16	Q98J17	Q98J17 rhizobium 1
18	68	6.6	171	17	O9UY30	O9UY30 pyrococcus
19	67.5	6.6	151	16	O8YK6	O8YK6 anabena sp
20	67.5	6.6	193	16	O9RXR1	O9RXR1 deinococcus
21	67	6.5	153	16	O92NX0	O92NX0 rhizobium m
22	67	6.5	181	16	O9UR36	O9UR36 neisseria m
23	67	6.5	192	2	Q9ZGB7	Q9ZGB7 streptomyce
24	67	6.5	194	16	O9RKK4	O9RKK4 streptomyce
25	67	6.5	197	17	O26850	O26850 methanobact
26	66.5	6.5	151	16	O9K3S9	O9K3S9 streptomyce
27	66.5	6.5	166	16	O9HWF2	O9HWF2 pseudomonas
28	66.5	6.5	187	16	O9CHE2	O9CHE2 lactococcus
29	66.5	6.5	197	6	O97673	O97673 sus scrofa
30	66	6.4	138	16	O9PD86	O9PD86 xylella fas
31	66	6.4	171	4	O96SA7	O96SA7 homo sapien
32	66	6.4	180	16	O91766	O91766 pseudomonas
33	65.5	6.4	147	17	O58707	O58707 pyrococcus
34	65	6.3	153	16	O9L099	O9L099 streptomyce
35	65	6.3	170	6	O28421	O28421 gorilla gor
36	65	6.3	178	17	O978C3	O978C3 thermoplas
37	65	6.3	188	2	O9Z5S9	O9Z5S9 zymomonas m
38	65	6.3	196	16	O8UB23	O8UB23 agrobacteri
39	64.5	6.3	189	16	O98FR8	O98FR8 rhizobium 1
40	64.5	6.3	190	17	O9HWA3	O9HWA3 halobacteri
41	64	6.2	129	2	O44423	O44423 agrobacteri
42	64	6.2	129	16	O8U753	O8U753 agrobacteri
43	64	6.2	131	5	O95Y12	O95Y12 bombyx mori
44	64	6.2	135	17	O8TXL9	O8TXL9 methanopyru
45	64	6.2	157	12	O93115	O93115 vaccinia vi

## ALIGNMENTS

## RESULT 1

ID	Q29124	PRELIMINARY;	PRT;	191 AA.
AC	Q29124;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	VLA-2 (Fragment).			
OS	Sue scrofa (pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.			
OX	NCBI_TaxID=96823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=ENDOTHELIAL CELLS;			
RX	MEDLINE=95036279; PubMed=7949129;			
RA	Bahou W.F., Potter C.L., Mirza H.;			
RT	"The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific			
RT	recognition sequence for endothelial cell attachment and spreading;			
RL	Blood 84:3734-3741(1994).			
DR	EMBL: Z12137; CAA78125.1; -			
DR	HSSP: P17301; IAOX.			
DR	InterPro: IPR002035; VWF_A.			
DR	Pfam: PF00092; VWA; 1.			
DR	PRINTS: PRO0453; VWFADOMAIN.			
DR	SMART: SM00327; VWA; 1.			
DR	PROSITE: PS50234; VWFA; 1.			
FT	NON_TER 1			
FT	NON_TER 191			
SO	SEQUENCE 191 AA; 20896 MW; 2E228B472EC699D8 CRC64;			

Query Match 32.8%; Score 336; DB 6; Length 191;

Best Local Similarity 43.9%; Pred. No. 2.4e-23;

Matches 69; Conservative 27; Mismatches 61; Indels 0; Gaps 0;

OY 42 MSEVOTFLRLRVKGLFDPEGIQVGLVYQGESPVHWSLGDFFRTKEEVYRAAKLSRRRG 101

DB 1 WDAVKNFLEKRVQGLDIGPTKTVGLIYANNPVVENLNTFKTKAEVWVATSTTQGG 60

QY 102 RETKTAQAIMVACTEGFSQSHGGRPARLLVVTTDGSHPDELPALKACEAGRTY 161  
 Db 61.DLTNTFKAIQYARDASVSAAGGRPCATKVMVVTTGSHDSGMLAAVVDCCNNINILRF 120  
 QY 162 GIAVIGHYLRORDPSSFLREIRTIASDPDERFFENV 198  
 Db 121 GIAVIGYLNRLMDTGNLIKIKAIKASIPTERYFFENV 157

## RESULT 2

ID 096930 PRELIMINARY; PRT; 146 AA.  
 AC 096930;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Thrombospondin-related anonymous protein (Fragment).  
 GN TRAP-2.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RA Doering C.D., Doering C.M.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X87840; CA61109.1;  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PSS0234; VWF\_A; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 146 AA; 16718 MW; A291357664B9F784 CRC64;

## Query Match

Best Local Similarity 23.9%; Score 92; DB 5; Length 146;  
 Pred. No. 0.7;  
 Matches 32; Conservative 29; Mismatches 49; Indels 24; Gaps 6;

QY 23 CPTVMDVIVLDGNSI--YPM-SEVQTLRLRLVGLFTDPEQIOVGL-----VQ 69  
 Db 11 CNYVDLTLLIDESASIGSKMKSHVLPPTDKIIXDITTSKBEVHGILLFSSKNRDVYT 70  
 QY 70 YGESPVHEWLDGPTKEVEVRAAKVLSRE--GRTKTAQAIMVACTEGFSQSHGGRPE 127  
 Db 71 YDELRVQ-----KDELKKVEKLDKDYCGGCTKILGALKIS--LENYTKHKNIIRD 121  
 QY 128 AARLLVVTGESH 141  
 Db 122 APKVTILFTDGNEN 135

## RESULT 3

ID 091415 PRELIMINARY; PRT; 194 AA.  
 AC 091415;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Beta 8 integrin (Fragment).  
 GN BETA 8 INTEGRIN.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRIN;  
 RX MEDLINE=95352850; PubMed=7542940;  
 RA Venstrom K., Reichardt L.;  
 FT "Beta 8 integrins mediate interactions of chick sensory neurons with

RT laminin-1, collagen IV, and fibronectin.";  
 RL Mol. Biol. Cell 6:419-431(1995).  
 DR EMBL; S78729; AAB34966.2; --  
 DR HSRF; P05106; IUV2.  
 DR InterPro; IPR002369; Integrin\_B.  
 DR Pfam; PF00362; Integrin\_B; 1.  
 DR PRINTS; PR01186; INTEGRINB.  
 DR ProDom; PD001811; Integrin\_B; 1.  
 DR SMART; SM00187; INB; 1.  
 KM Integrin.  
 FT NON\_TER 194  
 SQ SEQUENCE 194 AA; 21847 MW; B845301262330547 CRC64;

## Query Match

Best Local Similarity 36.1%; Score 74; DB 13; Length 194;  
 Pred. No. 48;  
 Matches 26; Conservative 7; Mismatches 27; Indels 12; Gaps 3;

QY 75 VHEWSLGD---FRTKEVEVRAAKVLSRREGRTKTAQAIMVACTEGFSQSH-GGRPDA 129  
 Db 67 IHVLSLTNIAEFRAVAVKQISGIDTPREGGFAMLQAAV-----QSHIGWRKDAK 119  
 QY 130 RLIVVVTGESH 141  
 Db 120 RLIVVVTGESH 131

## RESULT 4

ID 0920X7 PRELIMINARY; PRT; 134 AA.  
 AC 0920X7;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Von Willebrand factor (Fragment).  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.  
 OX NCBI\_TaxID=10141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HEART VENTRICLE;  
 RX MEDLINE=99345814; PubMed=10409222;  
 RA Preisig-Mueller R., Mederos y Schnitzler M., Daut J.;  
 RT "Separation of cardiomyocytes and coronary endothelial cells for cell-specific RT-PCR";  
 RL Am. J. Physiol. 277:H413-H416(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HEART VENTRICLE;  
 RA Preisig-Mueller R., Daut J., Mederos y Schnitzler M., Daut J.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF099069; AAD16289.1; --  
 DR HSRF; P04275; IAVQ.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR PROSITE; PSS0234; VWF\_A; 1.  
 FT NON\_TER 1  
 FT NON\_TER 134  
 SQ SEQUENCE 134 AA; 14852 MW; F6C97E931434B5F CRC64;

## Query Match

Best Local Similarity 22.9%; Score 72.5; DB 11; Length 134;  
 Pred. No. 40;  
 Matches 19; Conservative 19; Mismatches 22; Indels 23; Gaps 2;

QY 11 QPGSLAPTAQR-----CPTVMDVIVLDGNSI--YPMSEVQT 47  
 Db 28 QPGGLVVPPTAPASATPYLEDITELPMHDLVCSKLLDVLFDGSSKLEADFEVILKT 87  
 QY 48 FLRLVGLFTDPEQIOVGLVQY 70  
 Db 88 FVSVMERLHISQKRIRAVAVEY 110

```

RESULT 5
ID Q8TWC9 PRELIMINARY; PRT; 108 AA.
AC Q8TWC9;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE Nitrogen regulatory protein PII homolog.
GN MK1106.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Poluehin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natarle D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RT Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AB010400; AAM02319.1; -.
KW Complete proteome.
SQ SEQUENCE 108 AA; 12117 MW; 0678F55B81C69214 CRC64;

Query Match
Best Local Similarity 7.0%; Score 71.5; DB 17; Length 108;
Matches 25; Conservative 16; Mismatches 35; Indels 29; Gaps 5;

QY 55 KLFIDPEQ-----QVGL-----VOGESPVHMSLGDPR-TKEEVRAKXLSRE 100
DB 5 RLFVDPENGRVMMADAGVTGFAITETRGVAPPRMAGFELREDPESAIKLANDLSEK- 63
101 GRETTQAQIMVA-----CTEGFSQSHGGRPEARLLVVTVDGE 139
64 -----AVMIVTVPECEVEKLDMAERLAGEYTTIVDVE 100

RESULT 6
ID Q9RS46 PRELIMINARY; PRT; 164 AA.
AC Q9RS46;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein DR2281.
GN DR2281.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1.
RX MEDLINE=20036696; PubMed=10567266;
RA White O., Eison J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utechack T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Frazer C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RT Science 286:1571-1577(1999).
DR EMBL; AE002060; AAF1833.1; -.
DR TIGR; DR2281; -.
DR InterPro; IPR001159; DS_RBD.
DR Pfam; PF00035; dsrm; 1.

```

```

DR SMART; SM00358; DSRM; 1.
DR PROSITE; PS50137; DS_RBD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 164 AA; 17528 MW; 3055ABA9DAS587F CRC64;

Query Match
Best Local Similarity 6.9%; Score 71; DB 16; Length 164;
Matches 38; Conservative 19; Mismatches 66; Indels 28; Gaps 7;

QY 36 SNSIYPMSVQTFPLRIYGLKFLIDPEQIQVGLVGY-----GESPVH-----WSLG----- 81
DB 20 SRTLYPMNAKDDLRLV-----SLGLTPTFEAEGPAHRTTHVKVWSGGVIA 71
QY 82 --DERTKEEVRAAKNISRRGRETQTAQIMVACTEGFSQSHGGRPEARLLVVTVDGE 139
DB 72 TAEGRTKADARLAEALALRE-LDGSDAIPYPAPATPIAQOSEWPPIYAQVLAEAVAA 130
QY 140 SHDGEELPALKAC--EAGRTYRGIAVLGH 168
DB 131 MEFARE-DATLDEVNRDAGRFYRELADLGH 160

RESULT 7
ID Q58027 PRELIMINARY; PRT; 169 AA.
AC Q58027;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein PH0289.
GN PH0289.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyma A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RT DNA Res. 5:55-76(1998).
DR EMBL; AP000001; BAA29361.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 169 AA; 19349 MW; 3C5FB65945CEFF7A CRC64;

Query Match
Best Local Similarity 6.9%; Score 71; DB 17; Length 169;
Matches 35; Conservative 14; Mismatches 45; Indels 42; Gaps 6;

QY 40 YPMSEVQTFPLRIYGLKFLIDPEQIQVGLVGYGESP-----VHMSLGD----- 83
DB 52 YLSLHVSSAMRVLG-----VGLVQVKKRGRKAVFIATKQFSEWRSSAFYEKI 101
QY 84 -----RTKEEVRAAKNISRRGRETQTAQIMVACTEGFSQSHGGRPEARLLVVTVD 137
DB 102 LRDIDEFRENLRLAKELHEHKGSEVEIKKKLIALK-----RNEVARKLTLIMQ 153
QY 138 GESHDGEELPALKAC 153
DB 154 FRSE--BELTLVLEKC 167

RESULT 8
ID Q28796 PRELIMINARY; PRT; 170 AA.
AC Q28796;

```

DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Complement C4 (Fragment)  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.  
 NC NCB1\_Taxid=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA O'Huigin C.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92104634; PubMed=1729169;  
 RA Kawaguchi H., Zaleska-Ruczynska Z., Figueroa F., O'Huigin C.,  
 RA Klein U.;  
 RT "C4 genes of the chimpanzee, gorilla, and orang-utan: evidence for  
 RT extensive homogenization."  
 RL Immunogenetics 35:16-23(1992).  
 DR EMBL; Z31605; CAB3479.1; -.  
 DR HSSP; P01024; IC3D.  
 DR InterPro; IPR001559; Macrogloblina2.  
 DR Pfam; PF00207; AZM; 1.  
 FT NON\_TER 1 170 170  
 FT NON\_TER 1 170 170  
 SQ SEQUENCE 170 AA; 18045 MW; AE63A15A104B4854 CRC64;  
 Query Match 6.9%; Score 71; DB 6; Length 170;  
 Best Local Similarity 23.4%; Pred. No. 76;  
 Matches 33; Conservative 19; Mismatches 79; Indels 10; Gaps 4;  
 QY 34 DGSNSIYFVSEVOTFLRLVGLFIPEDQIQVGLVQYGSFVHEWEL-----GDPRTKE 87  
 DB 8 DGSYAWLSRDSSTWLTATVAVLVK-LSLADQVGVSEPKQETSMTLSQQQADSGSDS 66  
 QY 88 EVVRAAKNLSRREGRTKTAQA-IMVACTEGFS--OSHGGRPEARLLVVTDSHDE 144  
 DB 67 PVTHRGMGGLVGNDETVALTAFTVIALHHLGLAVFDQEGAEPLKQVEASISKANSFLGE 126  
 QY 145 ELPAALKACEAGRVTRYGIAV 165  
 DB 127 KASAGLLGAHAAAITAYALTL 147  
 RESULT 9  
 ID Q9X957 PRELIMINARY; PRT; 174 AA.  
 AC Q9X957;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical 19.0 kDa protein (Putative transcripional  
 DE regulator).  
 GN ORF4 OR SC02209 OR SCT087.04.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 NC NCB1\_Taxid=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=A;  
 RC STRAIN=A;  
 RA Vesterling S., Freese D., Engels A., Mohlleben W.;  
 RT "Analysis of the glnI region from Streptomyces coelicolor A3(2).";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);

RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 DR EMBL; Y13833; CAB43948.1; -.  
 DR EMBL; AL355752; CAB90857.1; -.  
 DR InterPro; IPR000835; HTH\_MARR.  
 DR PRINTS; PR00598; HTHMARR.  
 DR SMART; SM00347; HTH\_MARR; 1.  
 KW DNA-binding; Hypothetical protein; Transcription regulation.  
 SQ SEQUENCE 174 AA; 18985 MW; 58A15731E5033 CRC64;  
 Query Match 6.9%; Score 71; DB 16; Length 174;  
 Best Local Similarity 31.4%; Pred. No. 78;  
 Matches 27; Conservative 11; Mismatches 28; Indels 20; Gaps 4;  
 QY 118 FQSHGGRPEARLLVVTD-----GESHDGELPAALKACAG-----RTRRGIAVLG 167  
 DB 44 FAORNGMHPDVRALIALMDARAGEATTAGHGLGALGNSAGTALVRLERAG----- 98  
 QY 168 HYLRRQRPDSFLEIRITIASDPPER 193  
 DB 99 -HRRVRDE---RDRRVTVGVDER 119  
 RESULT 10  
 ID Q9M0V0 PRELIMINARY; PRT; 197 AA.  
 AC Q9M0V0;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical 21.8 kDa protein (MFDX1).  
 GN ATG05450 OR ATMPDX1.  
 OS Arabidopsis thaliana (mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NC NCB1\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vli D.M.,  
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,  
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dechta N.N.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA Takubo K., Nonaka Y., Mizutani S., Takenaka S., Takahashi M.,  
 RA Tsuyama S., Ohta D.,  
 RT "Identification and characterization of a mitochondrial electron  
 RT transfer chain in plants comprising of an adrenocortical ferredoxin  
 RT homologue and its oxidoreductase."  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Bahu J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
 RA Yamamura Y., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
 RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Full length cDNA of gene At4g05450 (GI:15235586)."  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Bahu J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Yamada K., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu S., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,  
 RA Davis R.W., Ecker J.R., Theologis A.,  
 RT "Arabidopsis Open Reading Frame (ORF) Clones."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP  
 RA EMBL: AL161503; CAB81087.1; -  
 RA EMBL: AB075738; BAB78226.1; -  
 RA EMBL: AY065280; AAL38756.1; -  
 RA DR EMBL: AY091324; AAM14263.1; -  
 RA DR HSSP: P00257; IAYF.  
 RA DR InterPro: IPR001055; Adrenodoxin.  
 RA DR InterPro: IPR000345; Cytochrome b5.  
 RA DR InterPro: IPR001041; Ferredoxin.  
 RA DR Pfam: PF00111; fer2; 1.  
 RA DR PRINTS: PR00355; ADRENODOXIN.  
 RA DR PROSITE: PS00814; ADX; 1.  
 RA DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN\_1.  
 RA KW Hypothetical protein.  
 RA SQ SEQUENCE 197 AA; 21830 MW; A35325C9F8C40A85 CRC64;  
 Query Match 6.9%; Score 71; DB 10; Length 197;  
 Best Local Similarity 22.7%; Pred. No. 93;  
 Matches 27; Conservative 15; Mismatches 39; Indels 38; Gaps 5;  
 Oy 94 KNLRSREG-----RETKTAQAIWA-----CTEGFSQSHGGRPEARLLVVVTGDS 138  
 Db 29 KNLHSYHYQLSLPVPFRQARTSQEAWFLKHKCTSTTSSENGDETEKTLIFVD- 87  
 Oy 139 ESHDEELPAALK-----ACEGRV--FGYGLVGLHYLRQRDS 177  
 Db 88 -KQDEELPVPKPIGMSVLEAHENDIDLEGACSLACSTCHVIMPTETYNKLEPT 144  
 RESULT 11  
 Q8ZME3 PRELIMINARY; PRT; 118 AA.  
 AC Q8ZME3;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Putative transcriptional regulators containing the CopG/Arc/MetJ  
 DE DNA-binding domain and a metal-binding domain.  
 GN STM2955;  
 OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OC NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=L72 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.,  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2."  
 RL Nature 413:852-856(2001).  
 DR EMBL: AE008835; AAL21835.1; -  
 DR InterPro: IPR005360; UPP0156.  
 DR Pfam: PF03693; UPP0156; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 118 AA; 13575 MW; 35280AFAE671DAD CRC64;  
 Query Match 6.9%; Score 70.5; DB 16; Length 118;  
 Best Local Similarity 21.4%; Pred. No. 52;  
 Matches 24; Conservative 15; Mismatches 24; Indels 49; Gaps 3;  
 Oy 81 GDFRKEEVRPAANKLSRREGRETKTAQAIWVACTEGFSQSHGGRPEARLLVVVTGDS 140  
 Db 47 GDFYQSVIRRESRLRLREKQESR-LQALRELLAEGL--NSEPQA----- 90  
 Oy 141 HDGEELPAALKACAGVTRGYGLVGLHYLRQRDPSSFLREITIASDPDE 192  
 Db 91 -----WEKDAFLRKVTGMKIPDE 109  
 RESULT 12  
 Q8Z448 PRELIMINARY; PRT; 118 AA.  
 AC Q8Z448;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Hypothetical protein STY3093.  
 GN STY3093.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OC NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
 RA Churcher S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahin A.W.,  
 RA Baker S., Baughman D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltham T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jajelski K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrett B.G.,  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18."  
 RL Nature 413:848-852(2001).  
 DR EMBL: AL627276; CAD06069.1; -  
 DR InterPro: IPR005360; UPP0156.  
 DR Pfam: PF03693; UPP0156; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 118 AA; 13606 MW; 43F36748B2164D59 CRC64;  
 Query Match 6.9%; Score 70.5; DB 16; Length 118;  
 Best Local Similarity 21.4%; Pred. No. 52;  
 Matches 24; Conservative 15; Mismatches 24; Indels 49; Gaps 3;  
 Oy 81 GDFRKEEVRPAANKLSRREGRETKTAQAIWVACTEGFSQSHGGRPEARLLVVVTGDS 140

Db 47 GYRTQSEVIRRESLRLREKQAESR-LOALRELLAGL---NSGEPOA----- 90  
 QY 141 HOGELPALAKACEAGRVTRYGIAVLGRDRDSSFLREIRTIASDPD 192  
 Db 91 -----WEXDAFLRKVKTKMIRKPE 109

## RESULT 13

08YHBI PRELIMINARY; PRT; 188 AA.

AC 08YHBI, 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)  
 DE Transcriptional regulator, TETR family.  
 GN BME10891.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;  
 RX MEDLINE=20020109; PubMed=11756688;  
 RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,  
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Resnik G.,  
 RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,  
 RA Selkov E., Elizer P.H., Hagius S., O'Callaghan D., Peterson J.-J.,  
 RA Haselkorn R., Kyprides N., Overbeek R.,  
 RT "The genome sequence of the facultative intracellular pathogen  
 RT Brucella melitensis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 DR EMBL: AE009529; AAL52072.1; -  
 DR InterPro: IPR001647; HTH\_Tetr.  
 DR Pfam: PF00440; Tetr\_1.  
 DR PRINTS: PR00455; HTHTETR.  
 DR Complete proteome.  
 KW SEQUENCE 188 AA; 20679 MW; B6F0102AD45104AA CRC64;

Query Match 6.8%; Score 70; DB 16; Length 188;  
 Best Local Similarity 25.0%; Pred. No. 1.1e+02;  
 Matches 28; Conservative 25; Mismatches 49; Indels 10; Gaps 5;

QY 85 TKEEVVRAKNSRREGRETKTAQAIMVACTEGFSQSGR-----PEARLLVVVDGSS 140  
 Db 9 SREKIRATTELAEQVGPAPHISIDA--VAARAGLSK--GGLXSPFTKAKLLEAMWEKYM 64  
 QY 141 HDGEELPALAKACEAGRVTRYGIAVLGHY-LRRQRPSSFLREIRTIASDPD 191  
 Db 65 QHEQMAVQETLQSGERNRVARAFDVYRIQADKEPPA-CGVLLALAEHPD 115

## RESULT 14

P71978 PRELIMINARY; PRT; 129 AA.

AC P71978, 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, last sequence update)  
 DE Hypothetical protein Rv1720c.  
 GN Rv1720C OR MTC04C12.05C OR MT1761.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigler K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jorgens K., Krogh A., Mclean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RT Nature 393:537-544(1998).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / OSHKOSH;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.,  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z81360; CAB03704.1; -  
 DR EMBL: AE007037; AAK46033.1; -  
 DR TIGR: MT1761; -  
 DR TubercuList; Rv1720C; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 129 AA; 13850 MW; D119173F733B06CD CRC64;

Query Match 6.7%; Score 69; DB 16; Length 129;  
 Best Local Similarity 23.0%; Pred. No. 81;  
 Matches 29; Conservative 19; Mismatches 50; Indels 28; Gaps 5;

QY 12 POGSLAPTAQR-----CPTVMDVIVLDGNSIYPMSEVOTFLRLVGLFTDPDQIQ 64  
 Db 17 PAG-AAVARRLRGETHVAPAHFDV-----EVIGAIRQAVVQLISDHEGL 60  
 QY 65 VGLVOYGSPPVHNSLGDFTKKEVVRANKLSRBEHETTAQAI--MACTEGFSQS 121  
 Db 61 VVVVNFSLPVRWPLKPTORAYQLRSTHTVA--DGAVALAGLGVLPTCGRLAQS 118  
 QY 122 HGRPE 127  
 Db 119 HGHNAE 124

## RESULT 15

O97P37 PRELIMINARY; PRT; 148 AA.

AC O97P37, 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)  
 DE Acetyltransferase, GNAI family.  
 GN SP1807.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tetelini H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Peterson J.D.,  
 RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 RL pneumoniae.";  
 RL Science 293:498-506(2001).  
 DR EMBL: AE007473; AAK75880.1; -  
 DR TIGR: SP1807; -





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:34:20 ; Search time 71 Seconds  
(without alignments)  
371.600 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337  
Perfect score: 1025  
Sequence: 1 GICARVDSFQGSGLAPTA.....FLREIRITASDPDERFFENV 198

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 698981

Minimum DB seq length: 0  
Maximum DB seq length: 198

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT:\*  
2: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1981.DAT:\*  
3: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1982.DAT:\*  
4: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1983.DAT:\*  
5: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1984.DAT:\*  
6: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1985.DAT:\*  
7: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1986.DAT:\*  
8: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1987.DAT:\*  
9: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1988.DAT:\*  
10: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1989.DAT:\*  
11: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1990.DAT:\*  
12: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1991.DAT:\*  
13: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1992.DAT:\*  
14: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1993.DAT:\*  
15: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1994.DAT:\*  
16: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1995.DAT:\*  
17: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1996.DAT:\*  
18: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1997.DAT:\*  
19: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1998.DAT:\*  
20: /SID82/gcgdata/geneseq/geneeqp-emb1/AA1999.DAT:\*  
21: /SID82/gcgdata/geneseq/geneeqp-emb1/AA2000.DAT:\*  
22: /SID82/gcgdata/geneseq/geneeqp-emb1/AA2001.DAT:\*  
23: /SID82/gcgdata/geneseq/geneeqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	912	89.0	195	23	AAU76853
2	912	89.0	195	23	AAU76862
3	492	48.0	193	23	AAU76854
4	492	48.0	193	23	AAU76863
5	417	40.7	195	23	AAU76851
6	417	40.7	195	23	AAU76860
7	389	38.0	195	23	AAU76852
8	389	38.0	195	23	AAU76861
9	374	36.5	185	22	AAU09125
10	285.5	27.9	148	22	AAU19634

11	285.5	27.9	148	22	AAU19794	Human novel extrac
12	285.5	27.9	148	23	ABP47854	Human polypeptide
13	285.5	27.9	148	23	ABP48014	Human polypeptide
14	282	27.5	103	22	AAU87675	Novel central nerv
15	282	27.5	103	22	AAU98822	Human novel extrac
16	282	27.5	103	23	ABP48042	Human polypeptide
17	244	23.6	176	23	ABP78816	Von Willebrand fac
18	171.5	16.7	192	23	AAU76855	Human integrin alp
19	171.5	16.7	192	23	AAU76864	Human integrin alp
20	149	14.5	191	23	AAU76850	Human integrin alp
21	149	14.5	191	23	AAU76859	Human integrin alp
22	148	14.4	180	23	ABP76376	Lymphocyte functio
23	148	14.4	184	21	AAE82349	Human CD11a I-doma
24	147	14.3	187	20	AAV21992	Human complement f
25	145	14.1	184	21	AAV21991	Rhesus CD11a I-dom
26	144	14.0	187	20	AAV21991	Human complement f
27	142	13.9	178	21	ABP08817	A von Willebrand A
28	140.5	13.7	185	22	AAE03654	Human extracellular
29	138	13.5	184	23	AAU76848	Human integrin alp
30	138	13.5	184	23	AAU76857	Human integrin alp
31	138	13.5	191	23	AAU76847	Human integrin alp
32	138	13.5	191	23	AAU76856	Human integrin alp
33	137	13.4	191	23	AAU76849	Human integrin alp
34	137	13.4	191	23	AAU76858	Human integrin alp
35	135	13.2	171	22	ABP27360	Human peptide #11
36	135	13.2	171	22	ABP32505	Peptide #11 encode
37	135	13.2	171	22	ABP18012	Peptide #11 encode
38	135	13.2	171	22	AAE53340	Human brain expres
39	135	13.2	171	22	AAE65717	Human bone marrow
40	135	13.2	171	22	AAE13577	Peptide #11 encode
41	135	13.2	171	22	AAE25974	Peptide #11 encode
42	135	13.2	171	22	AAE01329	Peptide #11 encode
43	135	13.2	171	22	ABG35347	Human peptide enco
44	135	13.2	172	22	ABP36981	Peptide #4487 enco
45	135	13.2	172	22	AAE17950	Peptide #4384 enco

## ALIGNMENTS

RESULT 1	AAU76853	
ID	AAU76853	standard; Protein: 195 AA.
XX	XX	
AC	AAU76853;	
XX	XX	
DT	21-MAY-2002	(first entry)
XX	XX	
DE	Human integrin alpha subunit Alpha 10 A domain.	
XX	XX	
KW	Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;	
KW	A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;	
KW	ischaemia-reperfusion injury; immune complex; parasitic disease.	
KW	antimicrobial; vasotropic; vulnerable; gene therapy.	
OS	Homo sapiens.	
XX	XX	
PN	WO200209737-A1.	
XX	XX	
PD	07-FEB-2002.	
XX	XX	
PF	31-JUL-2001; 2001WO-US33957.	
XX	XX	
PR	31-JUL-2000; 2000US-221950P.	
PR	11-JAN-2001; 2001US-0758493.	
PR	13-MAR-2001; 2001US-0805354.	
XX	XX	
PA	(GENO) GEN HOSPITAL CORP.	
XX	XX	
PI	Arnaut AM, Li R, Xiong J;	
XX	XX	
DR	WPI; 2002-188687/24.	
XX	XX	

PT Novel high affinity integrin polypeptide useful for treating restenosis  
PT and parasitic diseases, comprises all or part of variant integrin alpha  
PT subunit A domain or variant integrin beta subunit A-like domain  
XX  
PS Example 2; Fig 5; 55pp; English.  
XX  
CC The invention relates to a high affinity integrin polypeptide comprising  
CC all or part of a variant integrin alpha subunit A domain or a variant  
CC integrin beta subunit A-like domain. The polypeptide, preferably the  
CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by  
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
CC for determining if a test compound is a candidate compound for binding to  
CC CD1b or for treating an inflammatory disorder, by contacting a test  
CC compound with the polypeptide and determining if the test compound binds  
CC to the polypeptide. The integrin subunits are useful for reducing  
CC skeletal muscle injury, for treating disorders caused by  
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
CC diseases, to purify variant integrin polypeptide ligands and as bait  
CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
CC the human integrin alpha subunit Alpha 10 A domain.  
XX  
SQ Sequence 195 AA;  
Query Match 89.0%; Score 912; DB 23; Length 195;  
Best Local Similarity 100.0%; Pred. No. 6.5e-97;  
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLRLVGLFLIDPEQIQVGLVQGESPVHWSLGD 82  
DB 1 CPTVMDVIVLDGNSIYPMSEVQTFRLRLVGLFLIDPEQIQVGLVQGESPVHWSLGD 60  
QY 83 FRTKEVVAAPKXLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVTDSGSHD 142  
DB 61 FRTKEVVAAPKXLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVTDSGSHD 120  
QY 143 GEELPAAKACBAGVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 198  
DB 121 GEELPAAKACBAGVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 176  
RESULT 2  
AAU76862 standard; Protein; 195 AA.  
XX  
AC AAU76862;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human integrin alpha subunit Alpha 10 variant A domain.  
XX  
KW Human: integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;  
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;  
KW antiinflammatory; vasotropic; antiparasitic; vulnery; gene therapy;  
muten.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
XX FT Misc-difference 193 /note= "Wild-type Ile substituted by any other amino  
XX FT acid"  
XX FT  
XX PN WO200209737-A1.  
XX PD 07-FEB-2002.  
XX PF 31-JUL-2001; 2001WO-US23957.  
XX PR 31-JUL-2000; 2000US-221950P.  
PR 11-JAN-2001; 2001US-0758493.

PR 13-MAR-2001; 2001US-0805354.  
XX  
XX (GENO ) GEN HOSPITAL CORP.  
XX  
PI Arnaout AM, Li R, Xiong J;  
XX  
XX WPI; 2002-188687/24.  
XX  
PT Novel high affinity integrin polypeptide useful for treating restenosis  
PT and parasitic diseases, comprises all or part of variant integrin alpha  
PT subunit A domain or variant integrin beta subunit A-like domain  
XX  
PS Claim 53; Page -: 55pp; English.  
XX  
CC The invention relates to a high affinity integrin polypeptide comprising  
CC all or part of a variant integrin alpha subunit A domain or a variant  
CC integrin beta subunit A-like domain. The polypeptide, preferably the  
CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by  
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
CC for determining if a test compound is a candidate compound for binding to  
CC CD1b or for treating an inflammatory disorder, by contacting a test  
CC compound with the polypeptide and determining if the test compound binds  
CC to the polypeptide. The integrin subunits are useful for reducing  
CC skeletal muscle injury, for treating disorders caused by  
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
CC diseases, to purify variant integrin polypeptide ligands and as bait  
CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
CC a human integrin alpha subunit Alpha 10 variant A domain.  
CC Note: This variant sequence is not featured in the specification but has  
CC been derived from the wild-type protein shown in AAU76853.  
XX  
SQ Sequence 195 AA;  
Query Match 89.0%; Score 912; DB 23; Length 195;  
Best Local Similarity 100.0%; Pred. No. 6.5e-97;  
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLRLVGLFLIDPEQIQVGLVQGESPVHWSLGD 82  
DB 1 CPTVMDVIVLDGNSIYPMSEVQTFRLRLVGLFLIDPEQIQVGLVQGESPVHWSLGD 60  
QY 83 FRTKEVVAAPKXLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVTDSGSHD 142  
DB 61 FRTKEVVAAPKXLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVTDSGSHD 120  
QY 143 GEELPAAKACBAGVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 198  
DB 121 GEELPAAKACBAGVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 176  
RESULT 3  
AAU76854 standard; Protein; 193 AA.  
XX  
AC AAU76854;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human integrin alpha subunit Alpha 11 A domain.  
XX  
KW Human: integrin alpha subunit; A domain; Alpha 11; integrin beta subunit;  
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
KW ischaemia-reperfusion injury; immune complex; parasitic disease;  
KW antiinflammatory; vasotropic; antiparasitic; vulnery; gene therapy.  
XX  
OS Homo sapiens.  
OS  
XX  
XX WO200209737-A1.  
XX PD 07-FEB-2002.  
XX PF 31-JUL-2001; 2001WO-US23957.

XX 31-JUL-2000; 2000US-221950P.  
PR 11-JAN-2001; 2001US-0758493.  
PR 13-MAR-2001; 2001US-0805354.  
XX  
XX (GENO ) GEN HOSPITAL CORP.  
XX  
XX Arnaout AM, Li R, Xiong J;  
XX WPI; 2002-188687/24.  
XX  
XX Novel high affinity integrin polypeptide useful for treating restenosis  
XX and parasitic diseases, comprises all or part of variant integrin alpha  
XX subunit A domain or variant integrin beta subunit A-like domain  
XX  
XX Example 2; Fig 5; 55pp; English.  
XX  
XX The invention relates to a high affinity integrin polypeptide comprising  
XX all or part of a variant integrin alpha subunit A domain or a variant  
XX integrin beta subunit A-like domain. The polypeptide, preferably the  
XX CD11b alpha subunit A domain, where I at residue 332 has been replaced by  
XX C or A, F at residue 313 and A at residue 320 have been replaced by C, or  
XX V at residue 315 and A at residue 320 have been replaced by C, is useful  
XX for determining if a test compound is a candidate compound for binding to  
XX CD11b or for treating an inflammatory disorder, by contacting a test  
XX compound with the polypeptide and determining if the test compound binds  
XX to the polypeptide. The integrin subunits are useful for reducing  
XX skeletal muscle injury, for treating disorders caused by  
XX ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
XX diseases, to purify variant integrin polypeptide ligands and as bait  
XX proteins in two-hybrid or three-hybrid assays. This sequence represents  
XX the human integrin alpha subunit Alpha 11 A domain.  
XX  
XX Sequence 193 AA:  
SQ  
Query Match 48.0%; Score 492; DB 23; Length 193;  
Best Local Similarity 54.5%; Pred. No. 2.4e-48;  
Matches 96; Conservative 31; Mismatches 47; Indels 2; Gaps 1;  
QY 23 CPTWMDVIVLDGNSIYPMGEVOTFLRLVGLKLFIDPEQIOVGIVOGESPVHEWSLGD 82  
DB 1 COTYMDIYIVLDGNSIYPMGEVOTFLRLVGLKLFIDPEQIOVGIVOGESPVHEWSLGD 60  
QY 83 FRTKEEVRAAKNLRSREGRETQTAQIMVACTEGFSQSHGGRPPAARLLVVTDEGESH 142  
DB 61 YRSVDVVEAASHIQRGTEFRTAFGIEFARSEAFQK--GGRKAKKVMYIVITGESH 118  
QY 143 GEELPAALKACEAGVTRGYAVLGHYLRQRDPSSFLREIRTIASDDDERFFNV 198  
DB 119 SPDLKVIQOGERDVTRYAVAVLGVYNRGINPETFLNEIKYIASDPDKHFENV 174  
RESULT 4  
AAU76863  
ID AAU76863 standard; Protein; 193 AA.  
AC AAU76863;  
XX  
XX 21-MAY-2002 (first entry)  
XX  
XX Human integrin alpha subunit Alpha 11 variant A domain.  
XX  
XX Human; integrin alpha subunit; A domain; Alpha 11; integrin beta subunit;  
XX A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
XX ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;  
XX antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;  
XX mutein.  
XX  
XX Homo sapiens.  
XX OS Synthetic.  
XX  
XX Key Location/Qualifiers  
XX Misc-difference 191

FT /note= "wild-type Ile substituted by any other amino  
FT acid"  
PN W0200209737-A1.  
XX  
XX 07-FEB-2002.  
XX  
XX 31-JUL-2001; 2001WO-US223957.  
XX  
XX 31-JUL-2000; 2000US-221950P.  
PR 11-JAN-2001; 2001US-0758493.  
PR 13-MAR-2001; 2001US-0805354.  
XX  
XX (GENO ) GEN HOSPITAL CORP.  
XX  
XX Arnaout AM, Li R, Xiong J;  
XX WPI; 2002-188687/24.  
XX  
XX Novel high affinity integrin polypeptide useful for treating restenosis  
XX and parasitic diseases, comprises all or part of variant integrin alpha  
XX subunit A domain or variant integrin beta subunit A-like domain  
XX  
XX Claim 57; Page -: 55pp; English.  
XX  
XX The invention relates to a high affinity integrin polypeptide comprising  
XX all or part of a variant integrin alpha subunit A domain or a variant  
XX integrin beta subunit A-like domain. The polypeptide, preferably the  
XX CD11b alpha subunit A domain, where I at residue 332 has been replaced by  
XX C or A, F at residue 313 and A at residue 320 have been replaced by C, or  
XX V at residue 315 and A at residue 320 have been replaced by C, is useful  
XX for determining if a test compound is a candidate compound for binding to  
XX CD11b or for treating an inflammatory disorder, by contacting a test  
XX compound with the polypeptide and determining if the test compound binds  
XX to the polypeptide. The integrin subunits are useful for reducing  
XX skeletal muscle injury, for treating disorders caused by  
XX ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
XX diseases, to purify variant integrin polypeptide ligands and as bait  
XX proteins in two-hybrid or three-hybrid assays. This sequence represents  
XX a human integrin alpha subunit Alpha 11 variant A domain.  
XX Note: This variant sequence is not featured in the specification but has  
XX been derived from the wild-type protein shown in AAU76854.  
XX  
XX Sequence 193 AA:  
SQ  
Query Match 48.0%; Score 492; DB 23; Length 193;  
Best Local Similarity 54.5%; Pred. No. 2.4e-48;  
Matches 96; Conservative 31; Mismatches 47; Indels 2; Gaps 1;  
QY 23 CPTWMDVIVLDGNSIYPMGEVOTFLRLVGLKLFIDPEQIOVGIVOGESPVHEWSLGD 82  
DB 1 COTYMDIYIVLDGNSIYPMGEVOTFLRLVGLKLFIDPEQIOVGIVOGESPVHEWSLGD 60  
QY 83 FRTKEEVRAAKNLRSREGRETQTAQIMVACTEGFSQSHGGRPPAARLLVVTDEGESH 142  
DB 61 YRSVDVVEAASHIQRGTEFRTAFGIEFARSEAFQK--GGRKAKKVMYIVITGESH 118  
QY 143 GEELPAALKACEAGVTRGYAVLGHYLRQRDPSSFLREIRTIASDDDERFFNV 198  
DB 119 SPDLKVIQOGERDVTRYAVAVLGVYNRGINPETFLNEIKYIASDPDKHFENV 174  
RESULT 5  
AAU76851  
ID AAU76851 standard; Protein; 195 AA.  
AC AAU76851;  
XX  
XX 21-MAY-2002 (first entry)  
XX  
XX Human integrin alpha subunit Alpha 1 (CD49a) A domain.  
XX  
XX Human; integrin alpha subunit; A domain; CD49a; integrin beta subunit;

KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
 KW ischaemia-reperfusion injury; immune complex; parasitic disease; Alpha 1;  
 KW antiinflammatory; vasotropic; antiparasitic; vulnereary; gene therapy.  
 OS Homo sapiens.  
 XX WO200209737-A1.  
 PN 07-FEB-2002.  
 PD 31-JUL-2001; 2001WO-US23957.  
 PF 31-JUL-2000; 2000US-221950P.  
 PR 11-JAN-2001; 2001US-0758493.  
 PR 13-MAR-2001; 2001US-0805354.  
 XX (GHEO) GEN HOSPITAL CORP.  
 XX Arnaout AM, Li R, Xiong J;  
 PI WPI; 2002-188687/24.  
 DR Novel high affinity integrin polypeptide useful for treating restenosis  
 PT and parasitic diseases, comprises all or part of variant integrin alpha  
 PT subunit A domain or variant integrin beta subunit A-like domain -  
 XX Example 2; Fig 5; 55pp; English.  
 XX The invention relates to a high affinity integrin polypeptide comprising  
 CC all or part of a variant integrin alpha subunit A domain or a variant  
 CC integrin beta subunit A-like domain. The polypeptide, preferably the  
 CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by  
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
 CC for determining if a test compound is a candidate compound for binding to  
 CC CD1b or for treating an inflammatory disorder, by contacting a test  
 CC compound with the polypeptide and determining if the test compound binds  
 CC to the polypeptide. The integrin subunits are useful for reducing  
 CC skeletal muscle injury, for treating disorders caused by  
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
 CC diseases, to purify variant integrin polypeptide ligands and as bait  
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
 CC the human integrin alpha subunit Alpha 1 (CD49a) A domain.  
 CC Sequence 195 AA;  
 SQ  
 Query Match 40.7%; Score 417; DB 23; Length 195;  
 Best Local Similarity 45.5%; Pred. No. 1.1e-39;  
 Matches 80; Conservative 34; Mismatches 62; Indels 0; Gaps 0;  
 QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQVGLVQYGESEPVHEWSLGD 82  
 DB 1 CSTQLDIVIVLDGNSIYPMDSVTAFLNDLLKRMIDGPKQTVQVIGENVTHFNLNK 60  
 QY 83 FRTKEEVRAAKNLSRREGRETAKAIVACTEGFSGHGRPEAAALLVVTGDESHD 142  
 DB 61 YSSTEELVAAKKIVQGRQMTIALGTDTARKAFTEARCARGVKVMVIVTGDSEHD 120  
 QY 143 GEELPAALKACAGARTRYGIAVLGHYLRORDPSSEFLAEIRTIASDPDERFFENV 198  
 DB 121 NHRLLKVIQDCEDENLQRFSAIILGSGYRGNLSTEXFVEELKSIASEPTEKHFFNV 176  
 RESULT 6  
 AAU76860 standard; Protein; 195 AA.  
 XX AAU76860;  
 AC 21-MAY-2002 (first entry)  
 DT Human integrin alpha subunit Alpha 1 (CD49a) variant A domain.  
 XX

KW Human; integrin alpha subunit; A domain; CD49a; integrin beta subunit;  
 KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
 KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;  
 KW antiinflammatory; vasotropic; antiparasitic; vulnereary; gene therapy;  
 KW mutein; Alpha 1.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200209737-A1.  
 PN 07-FEB-2002.  
 PD 31-JUL-2001; 2001WO-US23957.  
 PF 31-JUL-2000; 2000US-221950P.  
 PR 11-JAN-2001; 2001US-0758493.  
 PR 13-MAR-2001; 2001US-0805354.  
 XX (GHEO) GEN HOSPITAL CORP.  
 XX Arnaout AM, Li R, Xiong J;  
 PI WPI; 2002-188687/24.  
 DR Novel high affinity integrin polypeptide useful for treating restenosis  
 PT and parasitic diseases, comprises all or part of variant integrin alpha  
 PT subunit A domain or variant integrin beta subunit A-like domain -  
 XX Claim 45; Page -; 55pp; English.  
 XX The invention relates to a high affinity integrin polypeptide comprising  
 CC all or part of a variant integrin alpha subunit A domain or a variant  
 CC integrin beta subunit A-like domain. The polypeptide, preferably the  
 CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by  
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
 CC for determining if a test compound is a candidate compound for binding to  
 CC CD1b or for treating an inflammatory disorder, by contacting a test  
 CC compound with the polypeptide and determining if the test compound binds  
 CC to the polypeptide. The integrin subunits are useful for reducing  
 CC skeletal muscle injury, for treating disorders caused by  
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
 CC diseases, to purify variant integrin polypeptide ligands and as bait  
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
 CC a human integrin alpha subunit Alpha 1 (CD49a) variant A domain.  
 CC Note: This variant sequence is not featured in the specification but has  
 CC been derived from the wild-type protein shown in AAU76851.  
 CC Sequence 195 AA;  
 SQ  
 Query Match 40.7%; Score 417; DB 23; Length 195;  
 Best Local Similarity 45.5%; Pred. No. 1.1e-39;  
 Matches 80; Conservative 34; Mismatches 62; Indels 0; Gaps 0;  
 QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQVGLVQYGESEPVHEWSLGD 82  
 DB 1 CSTQLDIVIVLDGNSIYPMDSVTAFLNDLLKRMIDGPKQTVQVIGENVTHFNLNK 60  
 QY 83 FRTKEEVRAAKNLSRREGRETAKAIVACTEGFSGHGRPEAAALLVVTGDESHD 142  
 DB 61 YSSTEELVAAKKIVQGRQMTIALGTDTARKAFTEARCARGVKVMVIVTGDSEHD 120  
 QY 143 GEELPAALKACAGARTRYGIAVLGHYLRORDPSSEFLAEIRTIASDPDERFFENV 198  
 DB 121 NHRLLKVIQDCEDENLQRFSAIILGSGYRGNLSTEXFVEELKSIASEPTEKHFFNV 176

RESULT 7  
AAU76852  
ID AAU76852 standard; Protein; 195 AA.  
XX  
AC AAU76852;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human integrin alpha subunit Alpha 2 (CD49b) A domain.  
XX  
KW Human; integrin alpha subunit; A domain; CD49b; integrin beta subunit;  
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
KW ischaemia-reperfusion injury; immune complex; parasitic disease; Alpha 2;  
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200209737-A1.  
XX  
PD 07-FEB-2002.  
XX  
PF 31-JUL-2001; 2001WO-US23957.  
XX  
PR 31-JUL-2000; 2000US-221950P.  
PR 11-JAN-2001; 2001US-0758493.  
PR 13-MAR-2001; 2001US-0805354.  
XX  
PA (GENO) GEN HOSPITAL CORP.  
XX  
PI Arnaout AM, Li R, Xiong J;  
XX  
DR WPI; 2002-188687/24.  
XX  
PT Novel high affinity integrin polypeptide useful for treating restenosis  
PT and parasitic diseases, comprises all or part of variant integrin alpha  
PT subunit A domain or variant integrin beta subunit A-like domain  
XX  
PS Example 2; Fig 5; 55pp; English.  
XX  
CC The invention relates to a high affinity integrin polypeptide comprising  
CC all or part of a variant integrin alpha subunit A domain or a variant  
CC integrin beta subunit A-like domain. The polypeptide, preferably the  
CC CD11b alpha subunit A domain, where I at residue 312 has been replaced by  
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
CC for determining if a test compound is a candidate compound for binding to  
CC CD11b or for treating an inflammatory disorder, by contacting a test  
CC compound with the polypeptide and determining if the test compound binds  
CC to the polypeptide. The integrin subunits are useful for reducing  
CC skeletal muscle injury, for treating disorders caused by  
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
CC diseases, to purify variant integrin polypeptide ligands and as bait  
CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
CC the human integrin alpha subunit Alpha 2 (CD49b) A domain.  
XX  
SQ Sequence 195 AA;  
XX  
Query Match 38.0%; Score 389; DB 23; Length 195;  
Best Local Similarity 44.9%; Pred. No. 2e-36; Indels 0; Gaps 0;  
Matches 79; Conservative 30; Mismatches 67;  
XX  
QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLVKGKLFIDPEQIOVGLVOYGSPPVHWSIGD 82  
DB 1 CPSLIDVVVVCDESISYPMQVAKNFKLEKFGVGLIGPKTQVGLIQVANNRRVFNINLT 60  
QY 83 FRTKEEVVRAAKNLSRRREGRTKTAQAIWVACTESFSGSHGGRPPAARLLVVVTDGESHD 142  
DB 61 YTKTEEMIVATQTSQYGGDLTNIFGALIOYARKVYVSAASGRRATKVMVVVTGDKSHD 120  
QY 143 GEELPAALKACAGAVTRVGIIVLGHVLRORDPSSFLREITTSDDPERFFNV 198  
DB 121 GMLKRAVIDQCNDHNIILRFGLIAGVGLNBNALDTKNLIKELKAIASIFTERYFENV 176

RESULT 8  
AAU76861  
ID AAU76861 standard; Protein; 195 AA.  
XX  
AC AAU76861;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human integrin alpha subunit Alpha 2 (CD49b) variant A domain.  
XX  
KW Human; integrin alpha subunit; A domain; CD49b; integrin beta subunit;  
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;  
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;  
KW murein; Alpha 2.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
Key Location/Qualifiers  
XX  
FT Misc-difference 193 /note= "Wild-type Ile substituted by any other amino  
FT acid"  
XX  
PN WO200209737-A1.  
XX  
PD 07-FEB-2002.  
XX  
PF 31-JUL-2001; 2001WO-US23957.  
XX  
PR 31-JUL-2000; 2000US-221950P.  
PR 11-JAN-2001; 2001US-0758493.  
PR 13-MAR-2001; 2001US-0805354.  
XX  
PA (GENO) GEN HOSPITAL CORP.  
XX  
PI Arnaout AM, Li R, Xiong J;  
XX  
DR WPI; 2002-188687/24.  
XX  
PT Novel high affinity integrin polypeptide useful for treating restenosis  
PT and parasitic diseases, comprises all or part of variant integrin alpha  
PT subunit A domain or variant integrin beta subunit A-like domain  
XX  
PS Claim 49; Page -; 55pp; English.  
XX  
CC The invention relates to a high affinity integrin polypeptide comprising  
CC all or part of a variant integrin alpha subunit A domain or a variant  
CC integrin beta subunit A-like domain. The polypeptide, preferably the  
CC CD11b alpha subunit A domain, where I at residue 312 has been replaced by  
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
CC for determining if a test compound is a candidate compound for binding to  
CC CD11b or for treating an inflammatory disorder, by contacting a test  
CC compound with the polypeptide and determining if the test compound binds  
CC to the polypeptide. The integrin subunits are useful for reducing  
CC skeletal muscle injury, for treating disorders caused by  
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
CC diseases, to purify variant integrin polypeptide ligands and as bait  
CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
CC a human integrin alpha subunit Alpha 2 (CD49b) variant A domain.  
CC Note: This variant sequence is not featured in the specification but has  
CC been derived from the wild-type protein shown in AAU76852.  
XX  
SQ Sequence 195 AA;  
XX  
Query Match 38.0%; Score 389; DB 23; Length 195;  
Best Local Similarity 44.9%; Pred. No. 2e-36; Indels 0; Gaps 0;  
Matches 79; Conservative 30; Mismatches 67;  
XX  
QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLVKGKLFIDPEQIOVGLVOYGSPPVHWSIGD 82  
DB 121 GMLKRAVIDQCNDHNIILRFGLIAGVGLNBNALDTKNLIKELKAIASIFTERYFENV 176

DB 1 CSRLIDVVVCDSESNISYPMDAVKNFLKFKVGLDIPGPTQVGLIQQVANNPREVENLNT 60  
QY 83 FRKKEVVAANKLRSREGRETKTAQAIWVACTEGFSQSGRPEARLLVVTDGESHSD 142  
DB 61 YKKEEMVIAVTSOTSYGGDLNTPGAIQYARKYVSAASGGRRSATKVMVVVTDGESHSD 120  
QY 143 GEELPALACACAGRTYRGIVLGHYLRORDPSSFLREIKTASDPDERFFNV 198  
121 GSWLKAVIQCNHNDILRFGIALVGLNNALDPTKLLIKEIKAIASIPTERYFNV 176  
RESULT 9  
AAU09125  
ID AAU09125 standard; protein, 185 AA.  
XX AAU09125;  
DT 16-JAN-2002 (first entry)  
XX Human integrin alpha2 I domain.  
DE  
XX Human; integrin alpha2; I domain; protein co-ordinate data; thrombolytic;  
KW cardiatic; cerebroprotective; hepatocytic; antiinflammatory; virucide;  
KW vasodilator; antiarteriosclerotic; thrombotic disorder;  
KW myocardial infarction; stroke; acute thrombosis; angioplasty;  
KW coronary bypass grafting; liver fibrosis; liver necrosis; hepatitis;  
KW arterial occlusion; restenosis; atherogenesis; anti-platelet therapy.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 17  
FT /label= OTHER  
FT /note= "Other= Unknown, represented by CPR in the  
FT specification"  
XX  
XX WO200173444-A2.  
XX  
XX 04-OCT-2001.  
XX  
XX 27-MAR-2001; 2001WO-GB01358.  
XX  
XX 28-MAR-2000; 2000US-192180P.  
XX  
XX (UYCA-) UNIV CAMBRIDGE TECH SERVICES LTD.  
XX  
XX Farndale RM, Emsley J, Knight CG, Barnes MJ, Liddington RC;  
XX  
XX WPI; 2001-648466/74.  
XX  
XX New methods for identifying inhibitors of an I-domain-containing  
PT polypeptide, particularly integrin, comprises use of coordinates of  
PT peptide/receptor crystal structure -  
XX  
XX Example; Page 47; 100pp; English.  
XX  
XX The invention relates to methods for identifying a potential inhibitor of  
CC an I-domain comprising polypeptide, particularly integrin, comprising  
CC using a three dimensional structure of the integrin alpha2 I-domain as  
CC defined in the specification to design or select a potential inhibitor.  
CC Identified inhibitors are used to treat a disorder or disease. Such  
CC disorders include thrombotic disorders, myocardial infarction and stroke,  
CC acute thrombosis associated with angioplasty and coronary bypass  
CC grafting, and with liver fibrosis or thrombotic complication of liver  
CC necrosis after hepatitis. Inhibition of platelet alpha2beta1 may be used  
CC to treat longer term occlusion of arteries, restenosis and atherogenesis.  
CC Collagen receptor antagonism may be used as anti-platelet therapy.  
CC The present sequence is the integrin alpha2 I domain.  
XX  
XX Sequence 185 AA;  
XX  
XX Query Match 36.5%; Score 374; DB 22; Length 185;  
XX Best Local Similarity 45.3%; Pred. No. 9.8e-35;

Matches 78; Conservative 28; Mismatches 66; Indels 0; Gaps 0;  
QY 27 MDVVVLGSGNSISYMSVQFLRLVGLKFLFIDPEQIOVGLVOGESPVHEWSLGDPTK 86  
DB 3 IDVVVVCBESNISTYKMDAVKNFLKFKVGLDIPGPTQVGLIQQVANNPREVFNLTYYTK 62  
QY 87 EEVVAANKLRSREGRETKTAQAIWVACTEGFSQSGRPEARLLVVTDGESHSDGSEL 146  
DB 63 EEMVIAVTSOTSYGGDLNTPGAIQYARKYVSAASGGRRSATKVMVVVTDGESHSDGSM 122  
QY 147 PAALACACAGRTYRGIVLGHYLRORDPSSFLREIKTASDPDERFFNV 198  
DB 123 KAVIQCNHNDILRFGIALVGLNNALDPTKLLIKEIKAIASIPTERYFNV 174  
RESULT 10  
AAU19634  
ID AAU19634 standard; Protein, 148 AA.  
XX AAU19634;  
AC  
XX  
DT 04-DEC-2001 (first entry)  
XX  
XX Human novel extracellular matrix protein, Seq ID No 284.  
DE  
XX Human; secreted extracellular matrix protein; immunomodulatory;  
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiatic; vascular;  
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cyostatic;  
KW antialzheimer's; immune/autoimmune disease; HIV infection; anaemia;  
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
KW cancer; hyperproliferative disorder; breast neoplasia; melanoma;  
KW Sezary syndrome; Gaucher's disease; neurological diseases;  
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
KW wound healing; immunogen; gene therapy; antisense; food additive.  
XX  
XX Homo sapiens.  
XX  
XX WO200155368-A1.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01348.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.



PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226868
PR	23-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0228294
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229519
PR	05-SEP-2000	2000US-0229530
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231412
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0233080
PR	12-SEP-2000	2000US-0233081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233663
PR	14-SEP-2000	2000US-0233664
PR	14-SEP-2000	2000US-0233670
PR	21-SEP-2000	2000US-0233655
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234597
PR	25-SEP-2000	2000US-0234598
PR	25-SEP-2000	2000US-0234599
PR	26-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0235837
PR	29-SEP-2000	2000US-0235867
PR	29-SEP-2000	2000US-0235868
PR	29-SEP-2000	2000US-0235869
PR	29-SEP-2000	2000US-0235870
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239355
PR	13-OCT-2000	2000US-0239357
PR	20-OCT-2000	2000US-0241826
PR	20-OCT-2000	2000US-0241827
PR	20-OCT-2000	2000US-0241828
PR	20-OCT-2000	2000US-0241875
PR	20-OCT-2000	2000US-0241876
PR	20-OCT-2000	2000US-0241877
PR	20-OCT-2000	2000US-0241878
PR	20-OCT-2000	2000US-0241879
PR	20-OCT-2000	2000US-0241880
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246619

[illegible]

Query Match 27.9%; Score 285.5; DB 22; Length 148;



PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465572/50.  
DR N-Psdb; AAS31365.  
XX  
XX  
PT Nucleic acid molecules encoding human secreted extracellular matrix  
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
PT Alzheimer's and Parkinson's diseases and cancers -  
PS  
PS Claim 11; SEQ ID No 444; 577bp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules encoding  
CC novel human secreted extracellular matrix proteins (SPs). The  
CC polynucleotides and proteins are used to prevent, treat a medical  
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
CC chickens or sheep. For example, disorders associated with decreased  
CC expression of SPs. The SP polynucleotide or a vector expressing them may  
CC be administered to treat diseases by gene therapy. Antisense molecules  
CC may be administered to down regulate expression of SPs by binding with  
CC the cells own genes and preventing their expression. The polynucleotides  
CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
CC used as antigens to produce antibodies and to identify modulators  
CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and  
CC antagonists may also be used to down regulate expression and activity of  
CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac  
CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
CC include wound healing, maintenance of organs before transplantation.

Query Match 27.9%; Score 285.5; DB 22; Length 148;

Best Local Similarity 51.3%; Pred. No. 1.2e-24;  
Matches 58; Conservative 19; Mismatches 35; Indels 1; Gaps 1;

Qy 1 GICARVDASFPQGSFLAFAQRCTYMDVIVLDGNSIYPSWSEVQFFLRLLVGNKLFIDP 60  
Db 21 GICSDVSPFPQVNSIAP-VQECSTQLDIVILVLDGNSIYPSWSDVTAFLNLLRRMDICP 79  
Qy 61 FOIQVGLVQGESFVHEMSLGDPRKKEVVAANKLSRERERKTAQAIWVA 113  
Db 80 KOTQVGIQVGENVTHEFNLNKYSTBEVLVAARKIVQRGGRQTMALGIDTA 132  
RESULT 12  
ABP47854  
ID ABP47854 standard; Protein, 148 AA.  
XX  
XX ABP47854;  
AC  
XX  
DT 23-AUG-2002 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 284.  
DE  
XX Human; noctropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antisticking; antianaemic; antirheumatic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; anti-inflammatory;  
KW antidiabetic; antidiabetic; antilicer; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.  
OS Homo sapiens.  
XX  
XX US2002042386-A1.  
XX  
XX 11-APR-2002.  
XX  
XX 17-JAN-2001; 2001US-0764870.  
XX  
XX 31-JAN-2000; 2000US-1790655P.  
XX 04-FEB-2000; 2000US-180628P.  
XX 28-JUN-2000; 2000US-214886P.  
XX 07-JUL-2000; 2000US-216647P.  
XX 07-JUL-2000; 2000US-216880P.  
XX 11-JUL-2000; 2000US-217487P.  
XX 11-JUL-2000; 2000US-217496P.  
XX 14-JUL-2000; 2000US-218290P.  
XX 26-JUL-2000; 2000US-220963P.  
XX 26-JUL-2000; 2000US-220964P.  
XX 14-AUG-2000; 2000US-224518P.  
XX 14-AUG-2000; 2000US-224519P.  
XX 14-AUG-2000; 2000US-225267P.  
XX 14-AUG-2000; 2000US-225268P.  
XX 14-AUG-2000; 2000US-225270P.  
XX 14-AUG-2000; 2000US-225447P.  
XX 14-AUG-2000; 2000US-225757P.  
XX 14-AUG-2000; 2000US-225758P.  
XX 22-AUG-2000; 2000US-226868P.  
XX 30-AUG-2000; 2000US-228924P.  
XX 01-SEP-2000; 2000US-229287P.  
XX 01-SEP-2000; 2000US-229343P.  
XX 01-SEP-2000; 2000US-229344P.  
XX 01-SEP-2000; 2000US-229345P.  
XX 05-SEP-2000; 2000US-229509P.  
XX 05-SEP-2000; 2000US-229513P.  
XX 08-SEP-2000; 2000US-231413P.  
XX 21-SEP-2000; 2000US-234223P.  
XX 21-SEP-2000; 2000US-234274P.  
XX 25-SEP-2000; 2000US-234977P.  
XX 27-SEP-2000; 2000US-235834P.  
XX 29-SEP-2000; 2000US-236327P.  
XX 29-SEP-2000; 2000US-236367P.  
XX 29-SEP-2000; 2000US-236367P.  
XX 29-SEP-2000; 2000US-236369P.  
XX 29-SEP-2000; 2000US-236370P.  
XX 02-OCT-2000; 2000US-236802P.  
XX 02-OCT-2000; 2000US-237037P.

PR 02-OCT-2000; 2000US-237038P.  
PR 02-OCT-2000; 2000US-237039P.  
PR 02-OCT-2000; 2000US-237040P.  
PR 13-OCT-2000; 2000US-239935P.  
PR 20-OCT-2000; 2000US-240960P.  
PR 20-OCT-2000; 2000US-241785P.  
PR 20-OCT-2000; 2000US-241809P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 17-NOV-2000; 2000US-249299P.  
PR 08-DEC-2000; 2000US-251866P.  
PR 08-DEC-2000; 2000US-251868P.  
PR 08-DEC-2000; 2000US-251869P.  
XX  
PA (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BARASH S C.  
XX  
PI Rosen CA, Ruben SM, Barash SC;  
XX  
PI WPI: 2002-470713/50.  
XX  
DR N-PSDB; AB066529.  
XX  
PT New nucleic acid encoding human proteins, useful for diagnosis,  
PT treatment and prevention of e.g. osteoporosis, also related  
PT polypeptides and antibodies  
XX  
PS Claim 11; SEQ ID NO 284; 235pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (AB066521-AB066785) and proteins  
CC (ABP47846-ABP4810) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=99909764870.  
XX  
SQ Sequence 148 AA;  
Query Match 27.9%; Score 285.5; DB 23; Length 148;  
Best Local Similarity 51.3%; Pred. No. 1,2e-24;  
Matches 58; Conservative 19; Mismatches 35; Indels 1; Gaps 1;  
QY 1 GTCARVDASFOGCSLAFTAQRCTYMDVVIIDGNSIYPMSEVQFLRLVKGKLFDP 60  
DB 21 GICSDVSPFQVYVNSIAP-VQECSTQDLIVIVDGSNSIYPMDSVTAFLNLLERMDIDP 79  
QY 61 EGIQVGLVQGSFVHMSLGDRTKEVVRAXKNSRREGRETAKQAQIMA 113  
DB 80 KOTQVGIVQYGVNTHFNLINKYSTEEVLVAKKIKVQRGRTMTALGIDTA 132  
RESULT 13  
ID ABP48014  
ID ABP48014 standard; Protein; 148 AA.  
XX  
XX ABP48014;  
AC  
XX  
XX  
DT 23-AUG-2002 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 444.  
XX  
XX Human; nocotropic; neuroprotective; cyostatic; dermatological; virucide;

KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antiskinking; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.  
OS Homo sapiens.  
XX  
XX US2002042386-A1.  
XX  
XX 11-APR-2002.  
XX  
PD 17-JAN-2001; 2001US-0764870.  
XX  
PF 31-JAN-2000; 2000US-179065P.  
XX  
PR 04-FEB-2000; 2000US-180628P.  
XX  
PR 28-JUN-2000; 2000US-214886P.  
XX  
PR 07-JUL-2000; 2000US-216647P.  
XX  
PR 07-JUL-2000; 2000US-216880P.  
XX  
PR 11-JUL-2000; 2000US-217487P.  
XX  
PR 11-JUL-2000; 2000US-217496P.  
XX  
PR 14-JUL-2000; 2000US-218290P.  
XX  
PR 26-JUL-2000; 2000US-220963P.  
XX  
PR 14-AUG-2000; 2000US-224518P.  
XX  
PR 14-AUG-2000; 2000US-224519P.  
XX  
PR 14-AUG-2000; 2000US-225267P.  
XX  
PR 14-AUG-2000; 2000US-225268P.  
XX  
PR 14-AUG-2000; 2000US-225270P.  
XX  
PR 14-AUG-2000; 2000US-225447P.  
XX  
PR 14-AUG-2000; 2000US-225757P.  
XX  
PR 14-AUG-2000; 2000US-225758P.  
XX  
PR 22-AUG-2000; 2000US-226688P.  
XX  
PR 30-AUG-2000; 2000US-228924P.  
XX  
PR 01-SEP-2000; 2000US-229287P.  
XX  
PR 01-SEP-2000; 2000US-229343P.  
XX  
PR 01-SEP-2000; 2000US-229344P.  
XX  
PR 01-SEP-2000; 2000US-229345P.  
XX  
PR 05-SEP-2000; 2000US-229509P.  
XX  
PR 05-SEP-2000; 2000US-229513P.  
XX  
PR 08-SEP-2000; 2000US-231413P.  
XX  
PR 21-SEP-2000; 2000US-234223P.  
XX  
PR 21-SEP-2000; 2000US-234274P.  
XX  
PR 25-SEP-2000; 2000US-234997P.  
XX  
PR 27-SEP-2000; 2000US-235834P.  
XX  
PR 29-SEP-2000; 2000US-236327P.  
XX  
PR 29-SEP-2000; 2000US-236367P.  
XX  
PR 29-SEP-2000; 2000US-236368P.  
XX  
PR 29-SEP-2000; 2000US-236369P.  
XX  
PR 29-SEP-2000; 2000US-236370P.  
XX  
PR 02-OCT-2000; 2000US-236802P.  
XX  
PR 02-OCT-2000; 2000US-237037P.  
XX  
PR 02-OCT-2000; 2000US-237038P.  
XX  
PR 02-OCT-2000; 2000US-237039P.  
XX  
PR 02-OCT-2000; 2000US-237040P.  
XX  
PR 13-OCT-2000; 2000US-239935P.  
XX  
PR 20-OCT-2000; 2000US-240960P.  
XX  
PR 20-OCT-2000; 2000US-241785P.  
XX  
PR 20-OCT-2000; 2000US-241809P.  
XX  
PR 01-NOV-2000; 2000US-244617P.  
XX  
PR 17-NOV-2000; 2000US-249299P.  
XX  
PR 08-DEC-2000; 2000US-251866P.  
XX  
PR 08-DEC-2000; 2000US-251868P.  
XX  
PR 08-DEC-2000; 2000US-251869P.  
XX  
XX  
XX (ROSE/) ROSEN C A.  
XX  
XX (RUBE/) RUBEN S M.  
XX  
XX (BARA/) BARASH S C.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
XX  
XX WPI: 2002-470713/50.

DR N-PSDB; ABO66689.  
XX New nucleic acid encoding human proteins, useful for diagnosis,  
PT treatment and prevention of e.g. osteoporosis, also related  
PR polypeptides and antibodies -  
XX  
PS Claim 11; SEQ ID NO 444; 235bp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABO66521-ABO66785) and proteins  
CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from USPTO at [seqdata.uspto.gov/sequence.html?docID=99909764870](http://seqdata.uspto.gov/sequence.html?docID=99909764870).  
XX  
SQ Sequence 148 AA:  
Query Match 27.9%; Score 285.5; DB 23; Length 148;  
Best Local Similarity 51.3%; Pred. No. 1.2e-24;  
Matches 58; Conservative 19; Mismatches 35; Indels 1; Gaps 1;  
OY 1 GICARVDSFOPGSLAPPAQRCPFYMDVIVLDGNSNITYPMSEVQTFRLVGLFIDP 60  
DB 21 GICSVSPFQVNSIAP-VQECSTQDLIVLDGNSNITYPMSTVTAFLINDLEMDIGP 79  
OY 61 EQLQVGLVQGESPVHEWSLGDFTKEEVYRAKNLSRREGRETYTAQAIWVA 113  
DB 80 KOTGVGIQVGENVTHFRNLNKYSTEEVLVAKKIVRGGRQMTALGIDTA 132  
RESULT 14  
AAU87675  
ID AAU87675 standard; Protein; 103 AA.  
XX  
AC AAU87675;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Novel central nervous system protein #585.  
XX  
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; lechemia; angiodogenesis;  
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
KW myocardial infarction; wound healing; cell proliferation; skin aging;  
KW food additive; food preservative; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200155318-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01332.  
XX  
PR 31-JAN-2000; 2000US-0178065.  
PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256119.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-581633/65.  
 DR N-PSDB; ABR44005.  
 XX  
 PT New isolated nucleic acid encoding a protein for diagnosing,  
 PT preventing, treating or ameliorating medical conditions and used as  
 PT food additives or preservatives -  
 XX

PS Claim 9; SEQ ID No 1193; 837bp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC novel central nervous system protein (I) and polypeptides (III) encoded  
 CC by (I), are used to treat a medical conditions and in diagnosis of a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and  
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi. Ocular disorders  
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
 CC adenocarcinomas and irritable bowel syndrome, reproductive system  
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
 CC acute kidney failure and blood related disorders e.g. myocardial  
 CC infarction. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,  
 CC

Query Match 27.5%; Score 282; DB 22; Length 103;  
 Best Local Similarity 57.1%; Pred. No. 1.8e-24;  
 Matches 56; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

QY 16 LAPTAORCPTMDVIVLDGNSISYPMSEVOTFRLRLGKLFIDPEQIQVGLVOYGESPV 75  
 DB 1 LCPHPGCGQTYMDIYIVLDGNSISYPMSEVOTFRLRLGKLFIDPEQIQVGLVOYGEDVV 60  
 QY 76 HEMSLGDRTRKEEYVRAAKNISREGRRTKTAQALMVA 113  
 DB 61 HEFLINDYRSVAKDYVEASHIEQRTETRTAFGIXFA 98

RESULT 15  
 ID AAU19822 standard; Protein; 103 AA.  
 XX  
 AC AAU19822;  
 XX  
 DT 06-DEC-2001 (first entry)  
 XX  
 DE Human novel extracellular matrix protein, Seq ID No 472.  
 XX  
 KW Human; secreted extracellular matrix protein; immunomodulatory;  
 KW Anti-HIV; antianemic; antithrombotic; antisclerotic; cardiac; vascular;  
 KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;  
 KW antialzheimers; immune/autoimmune disease; HIV infection; anaemia;  
 KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
 KW cancers; hyperproliferative disorder; breast neoplasia; melanoma;  
 KW Sezary syndrome; Gaucher's disease; neurological diseases;  
 KW Alzheimer's disease; Parkinson's disease; cardiovascular disorders;  
 KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
 KW wound healing; immunogen; gene therapy; antisense; food additive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO20015368-A1.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01348.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.

PR	17-MAR-2000	2000US-0190076
PR	18-APR-2000	2000US-0196123
PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209467
PR	28-JUN-2000	2000US-0209486
PR	30-JUN-2000	2000US-0215135
PR	07-JUL-2000	2000US-0216647
PR	07-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217487
PR	11-JUL-2000	2000US-0217496
PR	16-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226279
PR	18-AUG-2000	2000US-0226281
PR	22-AUG-2000	2000US-0226868
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227189
PR	30-AUG-2000	2000US-0228292
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229349
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231415
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233484
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233403
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234224
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235683
PR	29-SEP-2000	2000US-0235627
PR	29-SEP-2000	2000US-0235637
PR	29-SEP-2000	2000US-0235638
PR	29-SEP-2000	2000US-0235639
PR	29-SEP-2000	2000US-0235640
PR	29-SEP-2000	2000US-0235670
PR	02-OCT-2000	2000US-0235682
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239353

[illegible]

CC novel human secreted extracellular matrix proteins (SPs). The  
 CC polynucleotides and proteins are used to prevent, treat a medical  
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
 CC chickens or sheep. For example, disorders associated with decreased  
 CC expression of SPs. The SP polynucleotide or a vector expressing them may  
 CC be administered to treat diseases by gene therapy. Antisense molecules  
 CC may be administered to down regulate expression of SPs by binding with  
 CC the cells own genes and preventing their expression. The polynucleotides  
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
 CC used as antigens to produce antibodies and to identify modulators  
 CC (agonists and antagonists) of the SPs. The anti- (SP) antibodies and  
 CC antagonists may also be used to down regulate expression and activity of  
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
 CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac  
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
 CC include wound healing, maintenance of organs before transplantation,

Query Match 27.5%; Score 282; DB 22; Length 103;

Best Local Similarity 57.1%; Pred. No. 1.8e-24;

Matches 56; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

```
QY 16 LAPTAQRCPYMDVNVIVLDGNSIYPMSEVQTFRLVGLFIDPEQIOVGLVQGESPV 75
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 LCPHFGCOTMDIVIVLDGNSIYPMSEVQTFRLVGLFIDPEQIOVGLVQGESDVV 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 76 HEWSLGDFTKEVEVRANKLSRREGRETKTAQAINVA 113
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 HEFHLNDYRSVKQVVEASHIEGRGTEFRTAFGIKFA 98
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: July 16, 2003, 08:43:03  
 Job time : 72 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:45:06 ; Search time 53 Seconds  
(without alignments)  
443.669 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337  
Perfect score: 1025  
Sequence: 1 GICARVDSFQPGSLAPTA.....FLREIRTIASDPDERFFNNV 198

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues  
Total number of hits satisfying chosen parameters: 232099

Minimum DB seq length: 0  
Maximum DB seq length: 198

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA.\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	912	89.0	195	US-09-805-354-7	Sequence 7, Appli
2	912	89.0	195	US-10-144-259-7	Sequence 7, Appli
3	492	48.0	193	US-09-805-354-8	Sequence 8, Appli
4	492	48.0	193	US-10-144-259-8	Sequence 8, Appli
5	427	41.7	195	US-09-805-354-5	Sequence 5, Appli
6	427	41.7	195	US-10-144-259-5	Sequence 5, Appli
7	399	38.9	195	US-09-805-354-6	Sequence 6, Appli
8	399	38.9	195	US-10-144-259-6	Sequence 6, Appli
9	285.5	27.9	148	US-10-125-540-284	Sequence 284, App
10	285.5	27.9	148	US-10-125-540-444	Sequence 444, App
11	285.5	27.9	148	US-09-764-870-284	Sequence 284, App
12	285.5	27.9	148	US-09-764-870-444	Sequence 444, App
13	282	27.5	103	US-10-125-540-472	Sequence 472, App
14	282	27.5	103	US-09-764-870-472	Sequence 472, App
15	177.5	17.3	192	US-09-805-354-9	Sequence 9, Appli
16	177.5	17.3	192	US-10-144-259-9	Sequence 9, Appli
17	156	15.2	191	US-09-805-354-3	Sequence 3, Appli
18	156	15.2	191	US-10-144-259-3	Sequence 3, Appli
19	148	14.4	9	US-09-805-354-4	Sequence 4, Appli

20	148	14.4	184	9	US-10-144-259-4	Sequence 4, Appli
21	147	14.3	191	9	US-09-805-354-2	Sequence 2, Appli
22	147	14.3	191	9	US-10-144-259-2	Sequence 2, Appli
23	144	14.0	191	9	US-09-805-354-1	Sequence 1, Appli
24	144	14.0	191	9	US-10-144-259-1	Sequence 1, Appli
25	142	13.9	178	9	US-09-801-736A-4	Sequence 4, Appli
26	140.5	13.7	185	9	US-10-143-819-18	Sequence 18, Appli
27	135	13.2	171	10	US-09-864-761-33310	Sequence 33310, A
28	135	13.2	172	10	US-09-864-761-47637	Sequence 47637, A
29	123.5	12.0	193	9	US-09-984-245-293	Sequence 293, App
30	123.5	12.0	193	9	US-09-966-262-293	Sequence 293, App
31	123.5	12.0	193	9	US-09-983-966-293	Sequence 293, App
32	123.5	12.0	193	9	US-10-143-050-293	Sequence 293, App
33	111.5	10.9	160	9	US-09-801-736A-3	Sequence 3, Appli
34	91.5	8.9	159	9	US-09-864-761-49020	Sequence 49020, A
35	78	7.6	79	10	US-09-864-761-34986	Sequence 34986, A
36	73	7.1	196	9	US-10-156-761-10750	Sequence 10750, A
37	69	6.7	174	10	US-09-975-901-18	Sequence 18, Appli
38	69	6.7	181	10	US-09-975-901-5	Sequence 5, Appli
39	69	6.7	191	9	US-10-156-761-8978	Sequence 8978, Ap
40	67.5	6.6	124	9	US-09-796-692-1620	Sequence 1620, Ap
41	67.5	6.6	124	9	US-09-796-692-2496	Sequence 2496, Ap
42	67.5	6.6	124	9	US-10-040-862-1620	Sequence 1620, Ap
43	67.5	6.6	124	9	US-10-040-862-2496	Sequence 2496, Ap
44	67	6.5	54	10	US-09-864-761-49014	Sequence 49014, A
45	67	6.5	133	10	US-09-867-550-598	Sequence 598, App

## ALIGNMENTS

RESULT 1  
US-09-805-354-7  
Sequence 7, Application US/09805354  
Publication No. US20030078375A1  
GENERAL INFORMATION:  
APPLICANT: Arnaout, M. Amin  
APPLICANT: Li, Rui  
TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF  
FILE REFERENCE: 00786-536001  
CURRENT APPLICATION NUMBER: US/09/805,354  
CURRENT FILING DATE: 2002-06-04  
PRIOR APPLICATION NUMBER: US 09/758,493  
PRIOR FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: US 60/221,950  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 195  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-805-354-7

Query Match 89.0%; Score 912; DB 9; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1e-89;  
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 CPTMDVIVLDGNSIYPMSEVOTFLRLVYKLFIDBEQIOVGLVOYGSPVHEWSIGD 82  
DB 1 CPTMDVIVLDGNSIYPMSEVOTFLRLVYKLFIDBEQIOVGLVOYGSPVHEWSIGD 60  
QY 83 FRTSEVVRRAKNLSRRREGRTKTAQALMVACTGFSQSHGGRPAALLVVTGDESHD 142  
DB 61 FRTSEVVRRAKNLSRRREGRTKTAQALMVACTGFSQSHGGRPAALLVVTGDESHD 120  
QY 143 GEELPAALKAQEAQRVTRVYGLAVLGHYLRDRDSSFLREIRTIASDPDERFFNNV 198  
DB 121 GEELPAALKAQEAQRVTRVYGLAVLGHYLRDRDSSFLREIRTIASDPDERFFNNV 176

RESULT 2





```

: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PTL14C1
: CURRENT APPLICATION NUMBER: US/10/125,540
: PRIORITY FILING DATE: 2002-04-19
: Prior Application removed - See file wrapper or Palm
: NUMBER OF SEQ ID NOS: 646
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 284
: LENGTH: 148
: TYPE: PR1
: ORGANISM: Homo sapiens
: US-10-125-540-284

```

Query Match	27.9%;	Score 285.5;	DB 9;	Length 148;
Best Local Similarity	51.3%;	Pred. No. 9e-23;		
Matches 58;	Conservative 19;	Mismatches 35;	Indels 1;	Gaps 1

Qy 1 GICARVASTQPOGSLAPTRQPCRYMDVVIIVDGSNSIIPWSEVOTFRLRLVQKIFIP 60  
Db 21 GIGSDVPTFOVNSIAP-VQECSTQDVIIVIDGSNSIIPMDSVTAFINDLIERMDIGP 79

Qy 61 BOIIOVGIVQVGESEPVHEWISIGDPRTRKEVYVRAKNISRPREGPRTKAQAQIMWA 113  
Db 80 KQIVQVGIIVQGEVNTHEFNLKNTSSTBEVVAKKIIVQGRQGTMTALGIDTA 132

RESULT 10  
US-10-125

```

Sequence 444, Application US/10125540
Publication No. US2003005875A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PZ141
CURRENT APPLICATION NUMBER: US/10/125,540
CURRENT FILING DATE: 2002-04-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 646
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 444
LENGTH: 148
TYPE: PR1
ORGANISM: Homo sapiens
US-10-125-540-444

```

Query Match	27.9%	Score 285.5;	DB 9;	Length 148;
Best Local Similarity	51.3%;	Pred. No. 9e-23;		
Matches	58;	Conservative	19;	Mismatches 35;
			Indels	1;
			Gaps	1;

QY	DB
1 GICARVASFOPGSLAPFNORCPYADVIVIDGNSISIPMEVOTFRLKGFIDP	60
21 GICGSDFPTQVANSIAP-VQEGSTQDIYIVIDGNSISIPMDSVIAFLNDLLERNDICP	79
61 EIOYGVIVNGESPVEHMSLDFPTKEEVYRAKUNLSRGRGRTKQAQIAMA	113
80 KQDQVGVIVGQENVTHEFNINLKTSSTEEVIAAKKIVQGRGRTKALGIDTA	132

RESULT 11  
US-09-764

```

Sequence 284, Application US/09764870
Patent No. US20020042386A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT214
CURRENT APPLICATION NUMBER: US/09/764, 870
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 646
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 284
LENGTH: 148

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-284

```

Query Match	27.9%;	Score 285.5;	DB 10;	length 148;
Best Local Similarity	51.3%;	Pred. No. 9e-23;		
Matches 58;	Conservative 19;	Mismatches 35;	Indels 1;	Gaps 1

QY 1 GTCARVDASQPQGSGLAPTRKORCPYIMDVIVLVDGNSNITYPMEWEVOTFRLRLGKFIIP 60  
 DQ 21 GTCSDVASPTQVNNSIAP-VQECSTQLDIVIVLVDGNSNITYPMEWSVTAFLINDLDERMDIG 75  
 QY 61 EQLQVGLVYQGESPVHEWSLGDPRTEEEVYVRAAKNLSREGREGRTKAQAIWVA 113  
 DQ 80 KQIQVGLVYQGEVHTHEFNANKSSSTBEVYVVAAPKIVQIGQSGRTMTALGLDIT 132

RESULT 12

```

US-09-764-870-444
: Sequence 444, Application US/09764870
: Patent No. US20020042386A1
:
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.,
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P1214
: CURRENT APPLICATION NUMBER: US/09/764,870
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 646
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO: 444
: LENGTH: 148
: TYPE: PR1
: ORGANISM: Homo sapiens
US-09-764-870-444

```

Query Match	27.9%	Score 285.5;	DB 10;	Length 148;
Best Local Similarity	51.3%	Pred. No. 9e-23;		
Matches	58;	Conservative	19;	Mismatches 35;
				Indels 1;
				Gaps 1;

[illegible]

RESULT 13  
US-10-1.25

```

Sequence 472: Application US/10125540
Publication No. US20030055875A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT12C1
CURRENT APPLICATION NUMBER: US/10/125,540
CURRENT FILING DATE: 2002-04-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 646
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 472
LENGTH: 103

```

```

: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (96)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-125-540-472

Query Match      27.5%; Score 282; DB 9; Length 103;

```

Best Local Similarity 57.1%; Pred. No. 1.3e-22;  
Matches 56; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

Qy 16 LAPTORCPTVMDVYIVLDGNSIYPWSEVOTFLRLVGLKLFIDPEQIOVGLVOYGESPV 75  
Db 1 LCPHPPGQOTYMDIYIVLDGNSIYPWVEVOHFLNITLKRFYIGQIOVGVOYGEDV 60

Qy 76 HEWSIGDFRTKEEVVRAAKNLSRREGRETKTAQAIVVA 113  
Db 61 HEFHNDYRSVQDVVEAASHIEQGTETRTAFGIXFA 98

RESULT 14  
US-09-764-870-472  
; Sequence 472, Application US/09764870  
; Patent No. US20020042386A1  
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P1214  
; CURRENT APPLICATION NUMBER: US/09/764,870  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 646  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 472  
; LENGTH: 103  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (96)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-870-472

Query Match 27.5%; Score 282; DB 10; Length 103;  
Best Local Similarity 57.1%; Pred. No. 1.3e-22;  
Matches 56; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

Qy 16 LAPTORCPTVMDVYIVLDGNSIYPWSEVOTFLRLVGLKLFIDPEQIOVGLVOYGESPV 75  
Db 1 LCPHPPGQOTYMDIYIVLDGNSIYPWVEVOHFLNITLKRFYIGQIOVGVOYGEDV 60

Qy 76 HEWSIGDFRTKEEVVRAAKNLSRREGRETKTAQAIVVA 113  
Db 61 HEFHNDYRSVQDVVEAASHIEQGTETRTAFGIXFA 98

RESULT 15  
US-09-805-354-9  
; Sequence 9, Application US/09805354  
; Publication No. US20030078375A1  
; GENERAL INFORMATION:

; APPLICANT: Arnaout, M. Amin  
; APPLICANT: Li, Rui  
; APPLICANT: Xiong, Jian-Ping  
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 00786-536G01  
; CURRENT APPLICATION NUMBER: US/09/805,354  
; CURRENT FILING DATE: 2002-06-04  
; PRIOR APPLICATION NUMBER: US 09/758,493  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/221,950  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 192  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-805-354-9

Query Match 17.3%; Score 177.5; DB 9; Length 192;

Best Local Similarity 31.0%; Pred. No. 4.8e-11;  
Matches 54; Conservative 31; Mismatches 80; Indels 9; Gaps 5;

Qy 28 DWIVLDGNSIYP--MSEVOTFLRLVGLKLFIDPEQIOVGLVOYGESPVHEWSIGDFRT 85  
Db 6 EIALILDGSGSIDPPDPORAKDFTSNMKNRYEKCFCENFALVOYGVYOTEPFLRDSQD 65

Qy 86 KEEVVRRAKNSRREGRETKTAQAIVVACTEFSQSHGGRPEARLLVVTDGE-SHDGE 144  
Db 66 VMASLARVQNTI-QVGSVTKTASAMQVLDISFTSSGSRRAKSKVMVVLTDGGIFEDPL 124

Qy 145 ELPAALKACGKGVTRGIVLGHYLRQRPSSFLREIRTIASDPDERFFNV 198  
Db 125 NLITVINSPKMQGVVERFAIGV-GEFESKART---ARELNLIASDPDETHAFKV 173

Search completed: July 16, 2003, 08:51:47  
Job time : 54 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:41:56 / Search time 27 Seconds  
(without alignments)  
215.768 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025

Sequence: 1 GICARVDAFQPGSLAFTA.....FLREIRIYSDPDERFFENV 198

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 218932

Minimum DB seq length: 0

Maximum DB seq length: 198

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents, AA:  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	14.4	181	5	PCT-US96-01314-61
2	148	14.4	184	3	US-08-974-899-7
3	147	14.3	187	2	US-08-177-109A-62
4	147	14.3	187	2	US-08-687-706-62
5	147	14.3	187	5	PCT-US96-01314-60
6	145	14.1	184	3	US-08-974-899-8
7	144	14.0	187	2	US-08-177-109A-61
8	144	14.0	187	2	US-08-687-706-61
9	93	9.1	187	3	US-08-463-682-1
10	69	6.7	176	4	US-09-704-139-2
11	67.5	6.6	45	1	US-08-462-128-38
12	67.5	6.6	45	1	US-08-463-180-38
13	63.5	6.2	85	1	US-08-341-219-22
14	63.5	6.2	85	1	US-08-912-314A-22
15	60.5	5.9	75	4	US-08-864-357F-3
16	58.5	5.7	194	4	US-09-125-619-28
17	58	5.7	29	5	PCT-US96-01314-68
18	58	5.7	128	3	US-08-782-480-6
19	58	5.7	128	4	US-08-954-211-6
20	58	5.7	128	4	US-09-005-167A-6
21	57.5	5.6	128	4	US-09-176-741B-6
22	57.5	5.6	83	4	US-09-357-251-4
23	57.5	5.6	132	1	US-08-470-179-27
24	57.5	5.6	156	4	US-09-071-035-16
25	57.5	5.6	176	4	US-09-071-035-14
26	57.5	5.6	177	4	US-09-370-838-208
27	57.5	5.6	185	4	US-09-153-586-2

28	56.5	5.5	111	4	US-09-325-932A-63	Sequence 63, Appl
29	56.5	5.5	155	1	US-08-339-152A-22	Sequence 22, Appl
30	56.5	5.5	155	2	US-08-007-999B-7	Sequence 7, Appl
31	56.5	5.5	155	2	US-08-689-276A-7	Sequence 7, Appl
32	56	5.5	165	4	US-09-134-001C-5331	Sequence 5331, Ap
33	56	5.5	189	4	US-09-125-619-20	Sequence 20, Appl
34	56	5.5	195	1	US-08-438-753B-32	Sequence 32, Appl
35	56	5.5	195	1	US-08-443-883A-32	Sequence 32, Appl
36	56	5.5	195	2	US-08-631-328-32	Sequence 32, Appl
37	56	5.5	195	2	US-08-455-524B-32	Sequence 32, Appl
38	56	5.5	195	2	US-08-455-524B-32	Sequence 32, Appl
39	56	5.5	195	4	US-09-045-467-32	Sequence 32, Appl
40	55.5	5.4	46	1	US-08-462-128-40	Sequence 40, Appl
41	55.5	5.4	46	1	US-08-463-180-40	Sequence 40, Appl
42	55.5	5.4	99	2	US-08-795-444-1	Sequence 2, Appl
43	55.5	5.4	178	1	US-07-929-580B-2	Sequence 11, Appl
44	55.5	5.4	195	2	US-08-466-576B-11	Sequence 11, Appl
45	55.5	5.4	195	2	US-08-468-579B-11	Sequence 11, Appl

#### ALIGNMENTS

```
RESULT 1
PCT-US96-01314-61
; Sequence 61, Application PC/TUS9601314
; GENERAL INFORMATION:
; APPLICANT: M. Amin Arnaut
; TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
; TITLE OF INVENTION: ANTAGONISTS
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/01314
; FILING DATE: 30-JAN-96
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/380,167
; FILING DATE: 30-JAN-95
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00786/267001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 181 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-01314-61
```

Query Match 14.4%; Score 148; DB 5; Length 181;  
Best Local Similarity 25.4%; Pred. No. 1.5e-10;  
Matches 44; Conservative 37; Mismatches 70; Indels 22; Gaps 6;

QY 27 MDVIVIVDGSNIYTP--MSEVQTLRLRYLGLFDPEQIQGLVQYGESPHHMSLQGF- 83  
DB 6 VDLVFLFDGSMSLQPFDFQKILDFMKQVMKTL--SNISYQFAAVQFSTSYKTFDFDSYV 63

QY 84 ---RTKEVVRAAKNLSRREGRETKTAQATMVACTEGESGSHGGRPEARLLVVTDGESH 141  
DB 64 KMKDPALAKVYKHMILL-----TNTGAINVYATEVFREELGARPATKVLIIITDGEAT 119  
QY 142 DGEELPALKACEAGRVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERF 194  
DB 120 DSGNIDAKD-----IRYIIIGIGHGFOTKESQET-----LHKFASKPASEF 161

## RESULT 2

US-08-974-899-7  
Sequence 7, Application US/08974899  
Patent No. 6037454  
GENERAL INFORMATION:  
APPLICANT: Presta, Leonard G.  
TITLE OF INVENTION: Humanized Anti-CD11a Antibodies  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,899  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/031971  
FILING DATE: 11/27/96  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 40,378  
REFERENCE/DOCKET NUMBER: P1014R1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1994  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: Amino Acid  
TOPOLOGY: linear  
US-08-974-899-7

Query Match 14.4%, Score 148, DB 3, Length 184;  
Best Local Similarity 25.4%, Pred. No. 1.5e-10;  
Matches 44; Conservative 37; Mismatches 70; Indels 22; Gaps 6;

QY 27 MDVVIYVLDGNSIYF--WSEVOTFELRLVGLFIDEQIQVGLVQGESPVHEWSIGDP- 83  
DB 4 VDLVLFPSDSMSIOPDEFQKIIDFMKDVKKL--SNTSIQFAAVQFSYKTEFDSIV 61  
QY 84 --RTKEVVRAAKNLSRREGRETKTAQATMVACTEGESGSHGGRPEARLLVVTDGESH 141  
DB 62 KMKDPALAKVYKHMILL-----TNTGAINVYATEVFREELGARPATKVLIIITDGEAT 117  
QY 142 DGEELPALKACEAGRVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERF 194  
DB 118 DSGNIDAKD-----IRYIIIGIGHGFOTKESQET-----LHKFASKPASEF 159

RESULT 3  
US-08-177-109A-62  
Sequence 62, Application US/08177109A  
Patent No. 5869615

GENERAL INFORMATION:  
APPLICANT: Dennis E. Hourcade and Teresa J. Oglesby  
TITLE OF INVENTION: MODIFIED COMPLEMENT PROTEASES  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/177,109A  
FILING DATE: 03-JAN-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: WU 107  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-177-109A-62

Query Match 14.3%, Score 147, DB 2, Length 187;  
Best Local Similarity 26.7%, Pred. No. 2.1e-10;  
Matches 48; Conservative 36; Mismatches 84; Indels 12; Gaps 7;

QY 23 CPTY-MDVVIYVLDGNSIYF--WSEVOTFELRLVGLFIDEQIQVGLVQGESPVHEWS 79  
DB 1 CPPOQDIFVLDGSGSISRNFAFMNFVAVISO--FORP-STQFSIMQSNKFOETHFT 58  
QY 80 LGDRRTKEVVRAAKNLSRREGRETKTAQATMVACTEGESGSHGGRPEARLLVVTDGE 139  
DB 59 FEEFRRTSNPLSLASVHQLQG--FTYATRAIQNVVHRLFHASYGARRDARTKILVITDGK 117  
QY 140 SH-DGEELPALKACEAGRVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 198  
DB 118 KEGDSLQVYKQVIMPADAAGIRVAIGVLAQNR-----NSWKELNDIASKPSQEHIFKV 172

RESULT 4  
US-08-687-706-62  
Sequence 62, Application US/08687706  
Patent No. 5928892  
GENERAL INFORMATION:  
APPLICANT: Dennis E. Hourcade and Teresa J. Oglesby  
TITLE OF INVENTION: MODIFIED COMPLEMENT PROTEASES  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,706  
FILING DATE: 26-JUL-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/177,109  
FILING DATE: 03-JAN-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: WU 107 DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-687-706-62

Query Match 14.3%; Score 147; DB 2; Length 187;  
Best Local Similarity 26.7%; Pred. No. 2,1e-10;  
Matches 48; Conservative 36; Mismatches 84; Indels 12; Gaps 7;

Qy 23 CPTV-MDVIVLDGNSIYP--WSEVQFLRLVGKLFIDPEQIQVGLVQGESPVHMS 79  
Db 1 CPROQDIFPLIDSGSISRNFATMNFVRAVISQ-FQRP-STQFSLMQSNKFOQHT 58  
Qy 80 LGDPRTKEVYRAKNLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVVDGE 139  
Db 59 FEEFRRTSNPLSLASVHQLQG-FTYATPAIQNVVHRLFHASVYGRDARTILIVTDGK 117  
Qy 140 SH-DGEELPALAKACEAGRVTRYGIAYVGHYLRQRDPSSFLEIRTIASDPDERFFENV 198  
Db 118 KEGSLDYKDVIPMDADAGIIRYALGVGLAFQNR-----NSWKELNDIASRPSQEHIFKV 172

RESULT 5  
PCT-US96-01314-60  
Sequence 60, Application PC/TUS9601314  
GENERAL INFORMATION:  
APPLICANT: M. Amin Arnaout  
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN  
TITLE OF INVENTION: ANTAGONISTS  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01314  
FILING DATE: 30-JAN-96  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/380,167  
FILING DATE: 30-JAN-95  
ATTORNEY/AGENT INFORMATION:  
NAME: John W. Freeman  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 00786/267001

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-01314-60

Query Match 14.3%; Score 147; DB 5; Length 187;  
Best Local Similarity 26.7%; Pred. No. 2,1e-10;  
Matches 48; Conservative 36; Mismatches 84; Indels 12; Gaps 7;

Qy 23 CPTV-MDVIVLDGNSIYP--WSEVQFLRLVGKLFIDPEQIQVGLVQGESPVHMS 79  
Db 1 CPROQDIFPLIDSGSISRNFATMNFVRAVISQ-FQRP-STQFSLMQSNKFOQHT 58  
Qy 80 LGDPRTKEVYRAKNLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVVDGE 139  
Db 59 FEEFRRTSNPLSLASVHQLQG-FTYATPAIQNVVHRLFHASVYGRDARTILIVTDGK 117  
Qy 140 SH-DGEELPALAKACEAGRVTRYGIAYVGHYLRQRDPSSFLEIRTIASDPDERFFENV 198  
Db 118 KEGSLDYKDVIPMDADAGIIRYALGVGLAFQNR-----NSWKELNDIASRPSQEHIFKV 172

RESULT 6  
US-08-974-899-8  
Sequence 8, Application US/08974899  
Patent No. 6037454  
GENERAL INFORMATION:  
APPLICANT: Prestea, Leonard G.  
APPLICANT: Jardieu, Paula M.  
TITLE OF INVENTION: Humanized Anti-CD11a Antibodies  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,899  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/031971  
FILING DATE: 11/27/96  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 40,378  
REFERENCE/DOCKET NUMBER: P1014R1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1994  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-974-899-8

Query Match 14.1%; Score 145; DB 3; Length 184;



Sequence 1, Application US/08463682  
Patent No. 6008193  
GENERAL INFORMATION:  
APPLICANT: Leonard Garfinkel, et al.  
TITLE OF INVENTION: Cloning and Production of Human Von  
TITLE OF INVENTION: Willebrand Factor Gpib Binding Domain Polypeptides and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 1185 Avenue of Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,682  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 36537-B2-Y  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-463-682-1

Query Match  
Best Local Similarity 9.1%; Score 93; DB 3; Length 187;  
Matches 38; Conservative 46; Mismatches 72; Indels 28; Gaps 6;

Qy 23 CPTWMDVIVLDGNS--GIYPSWVOTFLRLVGLFDPEIQVGLVQYGESPVHMSL 80  
Db 1 CSRLDLVFLLDGSSRLSEAEFEVAKAFVDMMERLRISOKWVAVVEYHDSHAYIGL 60  
Qy 81 GDFRKEEVVRAKULSRREGRETQAQIMVACTEGFSQSHGR-PEARLLVVTGDSH 140  
Db 61 KDRKPSLRLRASQVKTAGSVASTSEVLTFLQITSKI--DPEPSRLILL- 113  
Qy 141 HDGEELPALAKACEGRVTRYGIIVLGHYLRQR-----DPSSFLREIRITIASD-PD 191  
Db 114 -----MASQEPQRSMRNFVRVY-QGLKKKQIVIVPGIGPHANLKQIRLIRKQABE 163  
Qy 192 ERF 195  
Db 164 NKA 167

RESULT 10  
US-09-704-139-2  
Sequence 2, Application US/09704139  
Patent No. 6420153  
GENERAL INFORMATION:  
APPLICANT: Kapellier, Rosanna  
TITLE OF INVENTION: 18232, A NOVEL DUAL SPECIFICITY PHOSPHATASE AND USES THEREFOR  
FILE REFERENCE: 10448-018001  
CURRENT APPLICATION NUMBER: US/09/704,139  
CURRENT FILING DATE: 2000-11-01

PRIOR APPLICATION NUMBER: US 60/185,772  
PRIOR FILING DATE: 2000-02-29  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 2  
LENGTH: 176  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1) (1390)  
OTHER INFORMATION: n = a, t, g, or c  
US-09-704-139-2

Query Match  
Best Local Similarity 6.7%; Score 69; DB 4; Length 176;  
Matches 37; Conservative 17; Mismatches 50; Indels 26; Gaps 7;

Qy 89 VRAAKNL---SRREGRTQAQIMVACTEGFSQSHGR-PEARLLVVTGDSHD- 142  
Db 19 LVRVAPSLFLGASABAQAEQARAGVTLGVNSRQGPAPGVAELRVVFPDPAEDL 78  
Qy 143 -----GEELPALKA-----CEAGRTRYGIIVLGHYLRQRDPS-----SFLREI 183  
Db 79 LAHLEPTCAAEAAVRAAGACLVYCKNGR--SRSAVCTAYLMHRGLSLAKAFQWVKA 136  
Qy 184 RTIASDPDERPF 195  
Db 137 RPVA-ERNPGFW 147

RESULT 11  
US-08-462-128-38  
Sequence 38, Application US/08462128  
Patent No. 5686059  
GENERAL INFORMATION:  
APPLICANT: Goetlinck, Paul F.  
TITLE OF INVENTION: Tondravi, Mehرداد  
TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lahive & Cockfield  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,128  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION:  
APPLICATION NUMBER: US 08/006,096  
FILING DATE: 15-JAN-1993  
APPLICATION NUMBER: US 07/866,403  
FILING DATE: 10-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: MGP-005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 amino acids  
TYPE: amino acid



COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/912,314A  
FILING DATE: 30-JUN-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/341,219  
FILING DATE: 05-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30742  
REFERENCE/DOCKET NUMBER: 8399-003-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 85 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-912-314A-22

Query Match 6.2%; Score 63.5; DB 4; Length 85;  
Best Local Similarity 32.9%; Pred. No. 2.2; Mismatches 26; Indels 19; Gaps 4;  
Matches 25; Conservative 6;

QY 110 IMVACTEFSQ--SHGRRPEARLLVVTGDSHDEELPALKACGAGRTYGIATIG<sup>167</sup>  
DB 14 LCVGQLSFQWHSWGYPGKREL-----DSFGTSEISEIKICEAGECG-----59  
QY 168 HYLRQRDPSSFLREI 183  
DB 60 -YLRPQR--RSLIRNI 72

RESULT 15  
US-08-864-357F-3  
Sequence 3, Application US/08864357F  
Patent No. 6255281  
GENERAL INFORMATION:  
APPLICANT: Claragen, Inc. & NIH  
TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato  
FILE REFERENCE: 116142/2  
CURRENT APPLICATION NUMBER: US/08/864,357F  
CURRENT FILING DATE: 1997-05-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3  
LENGTH: 75  
TYPE: PRT  
ORGANISM: rat  
US-08-864-357F-3

Query Match 5.9%; Score 60.5; DB 4; Length 75;  
Best Local Similarity 32.4%; Pred. No. 4.4; Mismatches 17; Indels 19; Gaps 5;  
Matches 24; Conservative 14;

QY 23 CPTWMDVV-IYLDGNSIY-----PMSEVO--TFLRLVGLKFLIDPEOIGVGLVQY 70  
DB 3 CPGLQVLEALLLGSESNYEALKFPNPASDLQNNAGTQKRLVDTL--PQTRINIVKL 59  
QY 71 GE---SPVHEWSL 80

Db 60 TEKILTSPLCEQDL 73

Search completed: July 16, 2003, 08:46:20  
Job time : 28 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:52 ; Search time 1.79677 Seconds  
(without alignments)  
1872.646 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192

Sequence: 1 YEHPHYGLPVGPGEFFKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	65.5	34.1	301	2	nodp protein - Azo
2	54	28.1	2139	2	vitellogenin - yel
3	52	27.1	101	2	hypothetical prote
4	52	27.1	437	1	transforming prote
5	52	27.1	564	2	gene gll protein -
6	52	27.1	649	2	C-eki protein FB27
7	52	27.1	713	2	C-eki protein FB28
8	52	27.1	728	1	transforming prote
9	52	27.1	750	2	C-eki protein FB29
10	51.5	26.8	390	2	hypothetical prote
11	51.5	26.8	507	1	cytochrome P450 52
12	51	26.6	402	2	hypothetical prote
13	51	26.6	890	2	NPAT transcription
14	51	26.6	1053	2	sucrose-phosphate
15	50.5	26.3	300	2	probable methylene
16	50.5	26.3	317	2	ATP sulfurylase, s
17	50	26.0	428	2	hypothetical prote
18	49.5	25.8	348	2	alcohol dehydrogen
19	49.5	25.8	666	2	glycosyl hydrolase
20	49	25.5	136	2	hypothetical prote
21	49	25.5	286	2	probable ribosomal
22	49	25.5	446	2	proline/betaine tr
23	49	25.5	753	2	dolichyl-phosphate
24	49	25.5	771	2	transposase - Yers
25	48.5	25.3	317	2	sulfate adenylyate
26	48.5	25.3	341	2	hypothetical prote
27	48.5	25.3	352	2	probable argc prot
28	48.5	25.3	505	2	probable tpr prote
29	48.5	25.3	653	2	probable tsbu prot

30	48.5	25.3	759	2	S53922	PM76 protein - yea
31	48.5	25.3	1046	2	B983324	probable rnd efflu
32	48.5	25.3	1066	2	AD29959	AcBr/AcrD/AcrF fam
33	48.5	25.3	1185	2	T36761	probable DNA polym
34	48	25.0	263	2	T35686	phosphatidylglycer
35	48	25.0	368	2	S57273	lignin peroxidase
36	48	25.0	492	2	G70749	probable hycr prot
37	48	25.0	661	1	TNBE12	74K alpha trans-in
38	48	25.0	728	2	S59964	procollagen-lysine
39	48	25.0	899	2	B75018	hypothetical prote
40	48	25.0	1262	2	T30524	protein phosphatas
41	48	25.0	1289	2	T18212	paraapoptal crystal
42	47.5	24.7	129	2	D70531	hypothetical prote
43	47.5	24.7	255	2	T13437	acid phosphatase h
44	47.5	24.7	299	1	ZZZRN	3'-phosphadenosin
45	47.5	24.7	299	2	G95932	probable sulfate a

## ALIGNMENTS

## RESULT 1

139754

nodp protein - Azospirillum brasiliense

C:Species: Azospirillum brasiliense

C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 11-May-2000

C:Accession: 139754

R:Vielle, C.; Elmerich, C.

Mol. Plant Microbe Interact. 3, 389-400, 1990

A:Title: Characterization of two Azospirillum brasiliense Sp7 plasmid genes homologous to

A:Reference number: 139754; MUID:92033082; PMID:2131098

A:Accession: 139754

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-301 <RES>

A:Cross-references: GB:M94886; NID:g142423; PIDN:AAA22185.1; PID:g142424

C:Genetics:

A:Gene: nodp

C:Superfamily: nodulation protein nodp

Query Match 34.1%; Score 65.5; DB 2; Length 301;

Best Local Similarity 48.4%; Pred. No. 0.11; 8; Indels 3; Gaps 2;

Matches 15; Conservative 5; Mismatches 5

QY 7 GTLPVPG--DEPKTLRVQNLGCVVSGLI 35

DB 228 GRPLNPGETEMR-RVRFRTLGCTPLSGAI 257

## RESULT 2

S46404

vitellogenin - yellow fever mosquito

C:Species: Aedes aegypti (yellow fever mosquito)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 21-Jul-2000

C:Accession: S46404

R:Chen, J.S.; Cho, W.L.; Raikhel, A.S.

J. Mol. Biol. 237, 641-647, 1994

A:Title: Analysis of mosquito vitellogenin cDNA. Similarity with vertebrate phosphatins a

A:Reference number: S46404; MUID:94210487; PMID:8158643

A:Accession: S46404

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2139 <CHE>

A:Cross-references: EMBL:U02548; NID:g472307; PIDN:AAA18221.1; PID:g472308

Query Match 28.1%; Score 54; DB 2; Length 2139;

Best Local Similarity 47.1%; Pred. No. 61;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 YEHPHYGLPVGPGEFFK 18

DB 366 DVHPDYKVPKGPADLK 382

## RESULT 3

S61172

hypothetical protein YDR377w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D9481.21

C:Species: Saccharomyces cerevisiae

C&gt;Date: 23-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 29-Oct-1999

C:Accession: S61172

R:Ring, H

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of S. cerevisiae cosmid 9481.

A:Reference number: S61159

A:Accession: S61172

A:Molecule type: DNA

A:Residues: 1-101 &lt;DIN&gt;

A:Cross-references: EMBL:U28373; NID:g849184; PIDN:AA64813.1; PID:g849198; MIPS:YDR377w

A:Experimental source: strain S288C (AB972)

C:Genetics:

A:Gene: SGD:ATP17

A:Cross-references: SGD:S0002785; MIPS:YDR377w

A:Map position: 4R

C:Superfamily: Saccharomyces hypothetical protein YDR377w

Query Match 27.1%; Score 52; DB 2; Length 101;

Best Local Similarity 47.6%; Pred. No. 4.1;

Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 3 VHPYGLPVGPGRPFRTIRV 23

DB 34 VHPYKSLPGCPAPAIKANTRL 54

## RESULT 4

TYFVSK

transforming protein ski - avian erythroblastosis virus (strain Sloan-Kettering)

C:Species: avian erythroblastosis virus

C&gt;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 02-Jun-1994

C:Accession: A32574

R:Stavner, E.; Brodeur, D.; Brennan, L.A.

Mol. Cell. Biol. 9, 4038-4045, 1989

A&gt;Title: The v-ski oncogene encodes a truncated set of c-ski coding exons with limited e

A:Reference number: A32574; MUID:89384634; PMID:2674685

A:Accession: A32574

A:Molecule type: DNA

A:Residues: 1-437 &lt;STA&gt;

C:Genetics:

A:Gene: v-ski

C:Superfamily: ski transforming protein

C:Keywords: transforming protein

Query Match 27.1%; Score 52; DB 1; Length 437;

Best Local Similarity 35.3%; Pred. No. 21;

Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGRPF-----KTLRVQNLGCVVSG 33

DB 46 PVMPGPFPMPSDRSTERCITLEGRTISCFVVG 79

## RESULT 5

gene gli protein - mouse (fragment)

C:Species: Mus sp. (mouse)

C&gt;Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 23-Feb-1997

C:Accession: I53106

R:Walterhouse, D.; Ahmed, M.; Slusarski, D.; Kalamaras, J.; Boucher, D.; Holmgren, R.; I

Dev. Dyn. 196, 91-102, 1993

A&gt;Title: gli, a zinc finger transcription factor and oncogene, is expressed during norma

A:Reference number: I53106; MUID:93372881; PMID:8364825

A:Accession: I53106

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues 1-564 &lt;RES&gt;

A:Cross-references: GB:S65038; NID:g410635; PID:g410636

C:Genetics:

A:Gene: gli

C:Superfamily: gli transforming protein

Query Match

Best Local Similarity 27.1%; Score 52; DB 2; Length 564;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 YEHPYGLPVGPGRP 15

DB 472 YEARGGSLPLGPGP 486

## RESULT 6

C32575

C-ski protein FB27 - chicken

C:Species: Gallus gallus (chicken)

C&gt;Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 18-Jun-1993

C:Accession: C32575

R:Sutrave, P.; Hughes, S.H.

Mol. Cell. Biol. 9, 4046-4051, 1989

A&gt;Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski ge

A:Reference number: A32575; MUID:89384635; PMID:2779576

A:Accession: C32575

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-649 &lt;SUT&gt;

A:Cross-references: GB:M28517

C:Superfamily: ski transforming protein

Query Match 27.1%; Score 52; DB 2; Length 649;

Best Local Similarity 35.3%; Pred. No. 32;

Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGRPF-----KTLRVQNLGCVVSG 33

DB 67 PVMPGPFPMPSDRSTERCITLEGRTISCFVVG 100

## RESULT 7

B32575

C-ski protein FB28 - chicken

C:Species: Gallus gallus (chicken)

C&gt;Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 18-Jun-1993

C:Accession: B32575

R:Sutrave, P.; Hughes, S.H.

Mol. Cell. Biol. 9, 4046-4051, 1989

A&gt;Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski ge

A:Reference number: A32575; MUID:89384635; PMID:2779576

A:Accession: B32575

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-713 &lt;SUT&gt;

A:Cross-references: GB:M28517

C:Superfamily: ski transforming protein

Query Match 27.1%; Score 52; DB 2; Length 713;

Best Local Similarity 35.3%; Pred. No. 35;

Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGRPF-----KTLRVQNLGCVVSG 33

DB 67 PVMPGPFPMPSDRSTERCITLEGRTISCFVVG 100

## RESULT 8

TYHUSK

transforming protein ski - human

C:Species: Homo sapiens (man)

C&gt;Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 18-Jun-1999

C:Accession: S06053

R:Nomura, N.; Sasamoto, S.; Ishii, S.; Date, T.; Matsui, M.; Ishizaki, R.



Nucleic Acids Res. 17, 5489-5500, 1989  
A>Title: Isolation of human cDNA clones of ski and the ski-related gene, smo.  
A:Reference number: S06052; MUID:89345144; PMID:2762147  
A:Accession: S06053  
A:Molecule type: mRNA  
A:Residues: 1-728 <NOM>  
A:Cross-references: GB:X15218; NID:936483; PIDN:CAA3288.1; PID:936484  
C:Comment: Transforming proteins ski may function as a transcription factor regulating m  
C:Genetics:  
A:Gene: GDB:SKI  
A:Cross-references: GDB:119595; OMIM:164780  
A:Map position: 1q22-1q24  
C:Superfamily: ski transforming protein  
C:Keywords: proto-oncogene; tandem repeat; transforming protein  
F:61-93/Region: proline-rich  
F:494-509/Region: serine-rich  
F:558-582,583-607,608-633/Region: 25-residue repeats

Query Match 27.1%; Score 52; DB 1; Length 728;  
Best Local Similarity 35.3%; Pred. No. 36;  
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPERF-----KTLRVONLGCYVSG 33  
DB 84 PVLPGPFPMPSDRSTERCTVLEGETISCFVVG 117

RESULT 9  
A32575  
C:ski protein FB29 - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 18-Jun-1993  
C:Accession: A32575  
R:Sutrave, P.; Hughes, S.H.  
Mol. Cell. Biol. 9, 4046-4051, 1989  
A>Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski ge  
A:Reference number: A32575; MUID:89384635; PMID:2779576  
A:Accession: A32575  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-750 <STU>  
A:Cross-references: GB:M28517  
C:Superfamily: ski transforming protein

Query Match 27.1%; Score 52; DB 2; Length 750;  
Best Local Similarity 35.3%; Pred. No. 37;  
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPERF-----KTLRVONLGCYVSG 33  
DB 67 PVMGPFPFMPDRSTERCTILEGETISCFVVG 100

RESULT 10  
T25996  
hypothetical protein ZK354.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25996  
R:Johnson, D.; Wamsley, P.; Bradshaw, H.  
submitted to the EMBL Data Library, February 1997  
A:Description: The sequence of C. elegans cosmid ZK354.  
A:Reference number: Z20120  
A:Accession: T25996  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-390 <JOH>  
A:Cross-references: EMBL:U88172; PIDN:ABA42257.1; GSPDB:GN00022; CESP:ZK354.2  
C:Genetics:  
A:Gene: CESP:ZK354.2  
A:Map position: 4  
A:introns: 59/3; 98/3; 175/2; 208/3; 372/3

Query Match 26.8%; Score 51.5; DB 2; Length 390;  
Best Local Similarity 54.5%; Pred. No. 21;  
Matches 12; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 11 VGPGEFFKTLRVONLGCYVVS 32  
DB 4 VGPGEFFKTLRVONLGCYVVS 24

RESULT 11  
S22973  
cytochrome P450 52A7 - yeast. (Candida tropicalis)  
N:Alternate names: cytochrome P450alk4  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Candida tropicalis  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Mar-2000  
C:Accession: S22973  
R:Seghezzi, W.; Meili, C.; Ruffner, R.; Kuenzi, R.; Sanglard, D.; Fiechter, A.  
submitted to the EMBL Data Library, June 1992  
A:Description: Isolation and characterization of additional members of the cytochrome P4  
A:Reference number: S22972  
A:Accession: S22973  
A:Molecule type: DNA  
A:Residues: 1-507 <SEG>  
A:Cross-references: EMBL:Z13011; NID:92661; PIDN:CAA78355.1; PID:92662  
C:Genetics:  
A:Gene: CYP52A7  
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxidoreductase  
F:303-478/Domain: cytochrome P450 homology <CYP>  
F:456/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 26.8%; Score 51.5; DB 1; Length 507;  
Best Local Similarity 40.7%; Pred. No. 29;  
Matches 11; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

QY 8 TLVPGPGEFFKTLRVONLGCYV-VSG 33  
DB 388 TLVPGGPGDGDPIVLRMGSIFISG 414

RESULT 12  
T37694  
hypothetical protein SPAC14C4.09 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
C:Accession: T37694  
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: Z21737  
A:Accession: T37694  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-402 <DEV>  
A:Cross-references: EMBL:Z98596; PIDN:CA11202.1; GSPDB:GN00066; SPDB:SPAC14C4.09  
A:Experimental source: strain 972h-; cosmid cl4C4  
C:Genetics:  
A:Gene: SPDB:SPAC14C4.09  
A:Map position: 1  
C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPBC646.06c

Query Match 26.6%; Score 51; DB 2; Length 402;  
Best Local Similarity 43.5%; Pred. No. 26;  
Matches 10; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 7 GLPVGPGPERFKTLRVONLGCY 29  
DB 374 GLASGCPPELVDSLSYNFAY 396

RESULT 13  
A48753

NFAT transcription factor phosphoprotein - mouse (fragment)  
 N:Alternate names: nuclear factor of activated T cells, phosphoprotein chain  
 C:Species: Mus musculus (house mouse)  
 C>Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 17-Mar-1999  
 C:Accession: A48753  
 R:McCaffrey, P.G.; Luo, C.; Karpola, T.K.; Jain, J.; Badalian, T.M.; Ho, A.M.; Burgeon, Science 262, 750-754, 1993  
 A:Title: Isolation of the cyclosporin-sensitive T cell transcription factor NFATp.  
 A:Reference number: A48753; PMID:94053710; PMID:8235597  
 A:Accession: A48753  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-890 <MCC>  
 A:Cross-references: GB:U02079  
 C:Keyword: phosphoprotein; transcription factor

Query Match 26.6%; Score 51; DB 2; Length 890;  
 Best Local Similarity 75.0%; Pred. No. 62;  
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 HPYGLPVGPGP 15  
 |||||  
 DB 805 HPGLCPVLPGR 816

RESULT 14  
 S34172  
 sucrose-phosphate synthase (EC 2.4.1.14) - potato  
 C:Species: Solanum tuberosum (potato)  
 C>Date: 22-Nov-1993 #sequence\_revision 03-Aug-1995 #text\_change 18-Jun-1999  
 C:Accession: S34172  
 R:Sonnwald, U.; Baener, A.  
 submitted to the EMBL Data Library, June 1993  
 A:Reference number: S34172  
 A:Accession: S34172  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1053 <SON>  
 A:Cross-references: EMBL:X73477; NID:g313264; PIDN:CAA51872.1; PID:g313265  
 C:Function:  
 A:Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc  
 A:Pathway: sucrose biosynthesis  
 C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology  
 C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis  
 F:168-651/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 26.6%; Score 51; DB 2; Length 1053;  
 Best Local Similarity 35.5%; Pred. No. 75;  
 Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEHPYGLPVGPGPEFTTLRVQNGCVV 31  
 ::|||::|||::|||::|||::|||  
 DB 911 FKVCKRGTVF--PSKELRKVMRIQALRCNAV 939

RESULT 15  
 T03464  
 probable methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) / methylenetetrahy  
 C:Species: Rhodobacter capsulatus  
 C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 03-Jun-2002  
 C:Accession: T03464  
 R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, U.; Haselkorn, R.; Fonstein, M.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997  
 A:Title: Sequence of a 188-kb segment of the chromosome of Rhodobacter capsulatus SB1003  
 A:Reference number: Z14955; PMID:97404404; PMID:9256491  
 A:Accession: T03464  
 A:Status: preliminary;  
 A:Molecule type: DNA  
 A:Residues: 1-300 <VIC>  
 A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16117.1; PID:g3128265  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

C:Keywords: hydrolase; multifunctional enzyme; oxidoreductase  
 F:9-287/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>

Query Match 26.3%; Score 50.5; DB 2; Length 300;  
 Best Local Similarity 38.5%; Pred. No. 22;  
 Matches 10; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

QY 2 EVHPYG---TLVPGPGPEFTTLRVQ 24  
 |||||::|||::|||::|||  
 DB 32 ETHPVGRVLVSIISIGPSPEIEVVRNQ 57

Search completed: July 16, 2003, 07:56:50  
 Job time : 4.79677 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:42 ; Search time 0.910689 Seconds  
(without alignments)  
1594.036 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192  
Sequence: 1 YEVHPYGLTPVGPGEFXTLRVQNLGCVVSLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	1167	1 ITNG_HUMAN	O75578 homo sapien
2	65.5	34.1	301	1 NODP_AZOBK	P28603 azospirillum
3	59	30.7	1189	1 ITNH_HUMAN	O94kx5 homo sapien
4	54.5	28.4	299	1 NODP_RHISB	O07308 rhizobium s
5	54	28.1	2148	1 VIRI_ABDAB	O16927 aedes aegypt
6	52	27.1	101	1 ATPK_YEAST	O06405 saccharomyc
7	52	27.1	348	1 SKI_MOUSE	O06098 mus musculu
8	52	27.1	437	1 SKI_AVIES	P17863 avian eryth
9	52	27.1	728	1 SKI_HUMAN	P13755 homo sapien
10	52	27.1	750	1 SKI_CHICK	P49140 gallus gall
11	52	27.1	1111	1 GLI1_MOUSE	P47806 mus musculu
12	51.5	26.8	507	1 CP5G_CANTR	P30609 candida tro
13	51	26.6	402	1 YD29_SCHPO	O13716 schizosacch
14	51	26.6	1053	1 SP5_SOLTU	O43845 solanum tub
15	51	26.6	1064	1 NPEC_MOUSE	O06091 mus musculu
16	50	26.0	489	1 YS45_CAEL	O09660 caenorhabdi
17	50	26.0	738	1 PLO3_HUMAN	O05058 homo sapien
18	49.5	25.8	815	1 GYRB_MYXXA	O33367 myxococcus
19	49	25.5	753	1 PMT3_YEAST	P47190 saccharomyc
20	49	25.5	1109	1 MSH7_ARATH	O95877 arabidopsi
21	48.5	25.3	352	1 ARGC_MYCTU	P94987 mycobacteri
22	48.5	25.3	653	1 YDE4_MYCTU	O11034 mycobacteri
23	48.5	25.3	759	1 PMT6_YEAST	P42934 saccharomyc
24	48	25.0	387	1 TGT_MOUSE	O91262 mus musculu
25	48	25.0	661	1 P102_VZVD	P03264 varicella-z
26	48	25.0	728	1 PLO1_RAT	O63321 rattus norv
27	48	25.0	1289	1 CSAB_BACUD	O45753 bacillus th
28	47.5	24.7	299	1 NODP_RHIME	P13441 rhizobium m
29	47.5	24.7	338	1 TRPD_RHOSH	O92748 rhodopacter
30	47.5	24.7	625	1 TROR_MOUSE	O08351 mus musculu
31	47.5	24.7	1065	1 KOGI_HUMAN	O75912 homo sapien
32	47	24.5	434	1 N075_LUPLU	O06841 lupinus lut
33	47	24.5	553	1 YF61_METUA	O58956 methanococc

ALIGNMENTS

RESULT 1	ID	ITNG_HUMAN	STANDARD	PRT	1167 AA.
AC	O75578; Q9UH28;	ITNG_HUMAN	STANDARD;	PRT;	1167 AA.
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	15-0UN-2002 (Rel. 41, Last annotation update)				
DE	Integrin alpha-10 precursor.				
GN	ITGA10.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Articular chondrocytes;				
RX	MEDLINE=98352078; PubMed=9685391;				
RA	Casper L., Hellman U., Lundgren-Akerlund E.,				
RT	"Isolation, cloning, and sequence analysis of the integrin subunit				
RT	alpha10, a beta1-associated collagen binding integrin expressed on				
RT	chondrocytes."				
RL	J. Biol. Chem. 273:20383-20389 (1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Heart, and Endothelial cells;				
RX	MEDLINE=20169197; PubMed=10702680;				
RA	Lehner K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,				
RT	Wang S.-X., Langley R., Krissansen G.W.;				
RT	"The integrin alpha10 subunit: expression pattern, partial gene				
RT	structure, and chromosomal localization."				
RL	Cytogenet. Cell Genet. 87:238-244 (1999).				
CC	- FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.				
CC	- SUBUNIT: HETERO DIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10				
CC	ASSOCIATES WITH BETA-1.				
CC	- SUBCELLULAR LOCATION: Type I membrane protein.				
CC	- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN				
CC	MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.				
CC	- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWF DOMAIN. INTEGRINS				
CC	WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.				
CC	- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 VWF DOMAIN.				
CC	- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.1sb-sib.ch/announce/">http://www.1sb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL; AF074015; AAC1952.1; -				
DR	EMBL; AF112345; AAF21944.1; -				
DR	EMBL; AF127273; AAF61638.1; -				
DR	HSSP; P17301; 1A0X.				
DR	Genew; HGNC:6135; ITGA10.				

```

OC Azospirillum.
OC NCBI_TaxId=192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp7;
RX MEDLINE=92033082; PubMed=2131098;
RA Vieille C., Elmerich C.;
RT "Characterization of two Azospirillum brasiliense Sp7 plasmid genes
RT homologous to Rhizobium meliloti nodQ.";
RL Mol. Plant Microbe Interact. 3:389-400(1990).
CC -1- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER
CC TO MOD FACTOR.
CC -1- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.
CC -1- SUBUNIT: SULFATE-ACTIVATING ENZYMES, NODP AND NODQ, MAY BE
CC PHYSICALLY ASSOCIATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/sib-ch).
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M94886; AAA22185.1; -
DR InterPro; IPR002500; PAPS_reduct.
DR Pfam; PF01507; PAPS_reduct; 1.
DR Kyr Plasmid; Nodulation; Transferrase; Nucleotidyltransferase.
SQ SEQUENCE 301 AA; 34137 MW; 8BD906E7B0237225 CRC64;

Query Match 34.1%; Score 65.5; DB 1; Length 301;
Best Local Similarity 46.4%; Pred. No. 0.075;
Matches 15; Conservative 5; Mismatches 8; Indels 3; Gaps 2.

Oy 7 GLEPVGPG--PEFKTLRVQNIAGYVWSGLI 35
Db 228 GRPLPNEGTEPEMR-RVRFRLGCTPLSGAI 257
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

RESULT 3
ITAH_HUMAN STANDARD; PRT; 1189 AA.
AC Q9UKX5; Q9UKQ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-11 precursor.
GN ITGAL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal heart, and Osteoblast;
RX MEDLINE=99417678; PubMed=10486209;
RA Lemett K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
RA Wang S.-X., Morris C.M., Kristiansen G.W.;
RT "Cloning, sequence analysis, and chromosomal localization of the novel
RT human integrin alpha11 subunit (ITGAL1).";
RL Genomics 60:179-187(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal muscle, and Uterus;
RX MEDLINE=99395147; PubMed=10464311;
RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;
RT "cDNA Cloning and Chromosomal Localization of Human alpha(11)
RT Integrin. A collagen-binding, 1 domain-containing, beta(1)-associated
RT integrin alpha-chain present in muscle tissues.";
RL J. Biol. Chem. 274:25735-25742(1999).
RN [3]
RP SEQUENCE OF 954-1188 FROM N.A.

```

CC TISSUE=Fibroblast; Andren N., Escivill X., Escarceller M., Sunoy L.;  
 RA Submitted (Jun-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND  
 CC HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO  
 CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO  
 CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS, IN  
 CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN  
 CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.  
 CC -1- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING  
 CC FETAL MUSCLE CELLS (IN VITRO).  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WFMA DOMAIN. INTEGRINS  
 CC WITH I-DOMAIN DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 WFMA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, AF109691; AAF01258.1; -  
 CC EMBL, AF137378; AAD51912.2; -  
 CC EMBL, AL359064; CAB94392.1; -  
 CC HSSP; P17301; IAOX.  
 CC Genew: HGNC:6136; ITG11.  
 CC MIM, 604789; -  
 CC InterPro; IPR000413; Integrin\_alpha.  
 CC InterPro; IPR002035; WFMA.  
 CC Pfam; PF00092; vwa; 1.  
 CC Pfam; PF01839; FG-GAP; 5.  
 CC PRINTS; PRO1185; INTEGRIN.  
 CC PRINTS; PRO0453; VMPADOMAIN.  
 CC SMART; SM00191; Int\_alpha; 5.  
 CC SMART; SM00327; vwa; 1.  
 CC PROSITE; PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.  
 CC PROSITE; PS50234; WFMA; 1.  
 CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 CC Signal; Repeat; Calcium; Magnesium; Polymorphism.  
 CC SIGNAL 1 22  
 CC CHAIN 23 1189 INTEGRIN\_ALPHA-11.  
 CC DOMAIN 23 1142 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 1143 1165 POTENTIAL.  
 CC DOMAIN 1166 1189 CYTOPLASMIC (POTENTIAL).  
 CC REPEAT 38 94 FG-GAP 1.  
 CC REPEAT 102 163 FG-GAP 2.  
 CC REPEAT 167 345 WFMA.  
 CC REPEAT 359 420 FG-GAP 3.  
 CC REPEAT 422 475 FG-GAP 4.  
 CC REPEAT 477 537 FG-GAP 5.  
 CC REPEAT 539 598 FG-GAP 6.  
 CC REPEAT 601 653 FG-GAP 7.  
 CC DOMAIN 1154 1162 POLY-LEU.  
 CC DOMAIN 1174 1177 POLY-ARG.  
 CC CA\_BIND 488 496 POTENTIAL.  
 CC CA\_BIND 551 559 POTENTIAL.  
 CC CA\_BIND 621 621 POTENTIAL.  
 CC DISULFID 76 83 BY SIMILARITY.  
 CC DISULFID 121 139 POTENTIAL.  
 CC DISULFID 129 159 POTENTIAL.  
 CC DISULFID 659 668 BY SIMILARITY.  
 CC DISULFID 674 729 BY SIMILARITY.  
 CC DISULFID 781 787 BY SIMILARITY.  
 CC DISULFID 881 893 BY SIMILARITY.  
 CC CARBOHYD 82 82 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 95 95 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 291 291 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 642 642 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 694 694 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 884 894 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 973 973 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1032 1032 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARIANT 433 433 V -> M.  
 FT VARIANT 524 524 R -> L.  
 FT VARIANT 524 524 R -> L.  
 FT VARIANT 972 972 L -> P.  
 FT VARIANT 972 972 L -> P.  
 FT VARIANT 1003 1003 I -> M.  
 FT VARIANT 1030 1030 /FTID=VAR\_009892.  
 FT VARIANT 1030 1030 MISSING.  
 FT VARIANT 1094 1094 /FTID=VAR\_009893.  
 FT VARIANT 1094 1094 L -> V.  
 FT VARIANT 1094 1094 L -> V.  
 SQ SEQUENCE 1189 AA, 133609 MW, 603030C08AA44C052 CRC64;  
 Query Match 30.7%; Score 59; DB 1; Length 1189;  
 Best Local Similarity 41.7%; Pred. No. 2.7;  
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
 Oy 12 GPGPEFTKTLRYONCGYVSSGLI 35  
 Db 981 GIGPFSCTFRIONLGLFPHGIM 1004  
 ID NODP\_RHISB STANDARD; PRT; 299 AA.  
 AC 007308;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4) (Sulfate adenylylate  
 DE transferase) (SAT) (ATP-sulfurylase small subunit) (Modulation protein  
 DE P).  
 GN NODP.  
 OS Rhizobium sp. (strain BR816).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBT\_Taxid=48291;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98083747; PubMed=9421916;  
 RA Laereans T., Cooliaet N., Verreth C., Snoeck C., Hellings N.,  
 RA Vanderleyden J., Martinez-Romero E.;  
 RT "Functional redundancy of genes for sulphate activation enzymes in  
 RL Rhizobium sp. BR816.";  
 RL Microbiology 143:3933-3942(1997).  
 CC -1- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER  
 CC TO NOD FACTOR.  
 CC -1- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.  
 CC -1- SUBUNIT: SULFATE-ACTIVATING ENZYMES, NODP AND NODD, MAY BE  
 CC PHYSICALLY ASSOCIATED (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@sb-sib.ch).
CC -----
CC DR EMBL: U59507; AAB95248.1; -.
CC DR InterPro: IPR002500; PAPS_reduct.
CC DR Pfam: PF01507; PAPS_reduct; 1.
CC KW Nucleotide transferase; Nucleotidyltransferase.
CC SQ SEQUENCE 299 AA; 34494 MW; 4F57C5F1983168F2 CRC64;

Query Match
Best Local Similarity 39.1%; Pred. No. 2 6; Length 299;
Matches 11; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

QY 9 LPVGPGE-FKTLRVQNLGCVVSGLI 35
Db 228 MPFEGEEVMDRVRFRITLGCYPLTGAI 255

RESULT 5
VITL_AEDAE STANDARD; PRT; 2148 AA.
AC Q16927; Q1692;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vitellogenin A1 precursor (VG) (PVG1) [Contains: Vitellin light chain
DE (VL); Vitellin heavy chain (VH)].
GN VGL.
OS Aedes aegypti (Yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Aedes.
OX NCBI_TaxID=7159;
[1]
RN RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=96035842; PubMed=7550249;
RA Romans P., Tu Z., Ke Z., Hagedorn H.H.;
RA "Analysis of a vitellogenin gene of the mosquito, Aedes aegypti and
RA comparisons to vitellogenins from other organisms.";
RL Insect Biochem. Mol. Biol. 25:939-958(1995).
RN [2]
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 17-24 AND 469-477.
RX STRAIN=UGALS; TISSUE=Fat body;
RX MEDLINE=94210487; PubMed=8158643;
RA Chen J.-S., Cho W.-L., Raikhel A.S.;
RA "Analysis of mosquito vitellogenin cDNA. Similarity with vertebrate
RA phospholipins and arthropod serum proteins.";
RL J. Mol. Biol. 237:641-647(1994).
RN [3]
RN RP SULFATION.
RX TISSUE=Fat body;
RX MEDLINE=90277688; PubMed=2351682;
RA Dhadiatala T.S., Raikhel A.S.;
RA "Biosynthesis of mosquito vitellogenin.";
RL J. Biol. Chem. 265:9924-9933(1990).
CC -1- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF
CC NUTRIENTS DURING EMBRYONIC DEVELOPMENT. MAY SUPPLY AROMATIC AMINO
CC ACIDS TO THE CUTICLE OF RAPIDLY DEVELOPING EMBRYOS.
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE FAT BODY, WHERE IT IS CLEAVED
CC IN THE ROUGH ENDOPLASMIC RETICULUM OR CIS-GOLGI BEFORE BEING
CC SECRETED INTO HEMOPLASM. IT IS THEN SEQUESTERED BY A SINGLE CLASS
CC OF RECEPTOR MEDIATED ENDOCYTOSIS IN THE OVARY.
CC -1- INDUCTION: SYNTHESIZED ONLY BY SEXUALLY MATURE FEMALE AFTER
CC INGESTION OF BLOOD.
CC -1- PTM: GLYCOSYLATED. PHOSPHORYLATED AND SULFATED. THE LARGE SUBUNIT
CC IS SULFATED MORE EXTENSIVELY THAN THE SMALL ONE.
CC -1- POLYMORPHISM: ALLELIC VARIATIONS DETECTED IN THE MOSQUITO
CC POPULATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC DR EMBL: L41842; AAB99486.1; -.
CC DR EMBL: U02348; AAA18221.1; -.
CC DR InterPro: IPR001747; Lipid_transprt_N.
CC DR InterPro: IPR001846; VWF_D.
CC DR Pfam: PF000094; vwd; 1.
CC DR Pfam: PF01347; Vitellogenin_N; 1.
CC DR SMART: SM00216; VWD; 1.
CC KW Glycoprotein; Phosphorylation; Storage protein; Signal; Polymorphism;
CC Sulfation.
FT SIGNAL 1 16
FT CHAIN 17 2148 VITELLOGENIN A1.
FT CHAIN 17 468 VITELLIN LIGHT CHAIN.
FT CHAIN 469 2148 VITELLIN HEAVY CHAIN.
FT MOD_RES 138 138 SULFATION (POTENTIAL).
FT MOD_RES 142 142 SULFATION (POTENTIAL).
FT MOD_RES 1046 1046 SULFATION (POTENTIAL).
FT MOD_RES 1049 1049 SULFATION (POTENTIAL).
FT MOD_RES 1053 1053 SULFATION (POTENTIAL).
FT MOD_RES 1542 1542 SULFATION (POTENTIAL).
FT MOD_RES 1543 1543 SULFATION (POTENTIAL).
FT MOD_RES 1549 1549 SULFATION (POTENTIAL).
FT MOD_RES 1716 1716 SULFATION (POTENTIAL).
FT MOD_RES 1785 1785 SULFATION (POTENTIAL).
FT MOD_RES 1788 1788 SULFATION (POTENTIAL).
FT MOD_RES 1801 1801 SULFATION (POTENTIAL).
FT MOD_RES 1803 1803 SULFATION (POTENTIAL).
FT MOD_RES 1867 1867 SULFATION (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1119 1119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1315 1315 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1631 1631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1675 1675 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1956 1956 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 417 440 POLY-SER.
FT DOMAIN 496 512 POLY-SER.
FT DOMAIN 520 548 POLY-SER.
FT DOMAIN 2007 2025 POLY-SER.
FT DOMAIN 2033 2042 POLY-SER.
FT CONFLICT 39 47 MISSING (IN REF. 2).
FT CONFLICT 906 906 R -> S (IN REF. 2).
FT CONFLICT 1367 1367 F -> Y (IN REF. 2).
FT CONFLICT 1576 1576 C -> Y (IN REF. 2).
FT CONFLICT 1757 1757 N -> Y (IN REF. 2).
FT CONFLICT 1935 1935 A -> P (IN REF. 2).
FT CONFLICT 2031 2031 G -> S (IN REF. 2).
SQ SEQUENCE 2148 AA; 250249 MW; 2AB4DF63AC766AF CRC64;

Query Match
Best Local Similarity 28.1%; Score 54; DB 1; Length 2148;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EVHPGTLVPVGPGEF 18
Db 375 DVHPYDKVMPGPAEDLK 391

RESULT 6
ATPK_YEAST STANDARD; PRT; 101 AA.
AC Q06405;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase F chain, mitochondrial precursor (EC 3.6.3.14).
GN ATP17 OR YDR317W OR D9481.21.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC STRAIN=D273-10B/A/H/U;
RX MEDLINE=97433297; PubMed=9288937;
RA Spanghel C., Valler J., Arselin G., Graves P.-V., Velours J.;
RT "The subunit F of mitochondrial yeast ATP synthase -- characterization
RT of the protein and disruption of the structural gene ATP17."
RL Eur. J. Biochem. 247:1111-1117(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fallon L., Galtung S., Greco T., Kirsten J.,
RA Kueba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Lacroille P., Le T.,
RA Martin E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevasaki E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DDBJ databases.
CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC
CC COMPONENT (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC -1- CATALYTIC ACTIVITY: ATP + H(2O) + H(+) (1n) = ADP + phosphate +
CC H(+) (Out).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. IN YEAST, THE
CC DIMERIC FORM OF ATP SYNTHASE CONSISTS OF 18 POLYPEPTIDES: ALPHA,
CC BETA, GAMMA, DELTA, EPSILON, 4 (B), 5 (OSCP), 6 (A), 8, 9 (C), D,
CC E (TIM11), F, G, H, I, J AND K.
CC -1- MASS SPECTROMETRY: MW=10565; MW_ERR=2; METHOD=Electrospray;
CC RANGE=7-101.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U72652; AAB70108.1; -
DR EMBL: U28373; AAB64813.1; -
DR SGD: S0002785; ATP17.
KW ATP synthase; Hydrogen ion transport; CF(0); Mitochondrion;
KW Transmembrane protein;
FT TRANSIT 1 6 MITOCHONDRION.
FT CHAIN 7 101 ATP SYNTHASE F CHAIN.
SQ SEQUENCE 101 AA; 11312 MW; 3DEF593AE435551 CRC64;

```

```

Query Match 27.1%; Score 52; DB 1; Length 101;
Best Local Similarity 47.6%; Pred. No. 1.9;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

```

```

OY 3 VHPVGLFVGPGEFFKTLRV 23
   |||||
DB 34 VHFVSLPGPAPPAIKANRL 54

```

```

RESULT 7
SKI_MOUSE STANDARD; PRT; 348 AA.
AC 060698;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Ski oncogene (C-ski) (Fragment).
GN SKI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/J;
RX MEDLINE=96127473; PubMed=8573720;
RA Nanciu S., Lyons G.E., Micales B.K., Heyman H.-C., Colmenares C.,
RA Stavezer E.;
RT "Enhanced expression of mouse c-ski accompanies terminal skeletal
RT muscle differentiation in vivo and in vitro."
RL Dev. Dyn. 204:291-300(1995).
CC -1- FUNCTION: MAY PLAY A ROLE IN TERMINAL DIFFERENTIATION OF SKELETAL
CC MUSCLE CELLS BUT NOT IN THE DETERMINATION OF CELLS TO THE MYOGENIC
CC LINEAGE.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: IS EXPRESSED IN AN UNIFORM PATTERN IN ALL
CC EMBRYONIC CELLS PRIOR TO SKELETAL MUSCLE CELL FORMATION IN THE
CC MYOTOMES OF SOMITES. EXPRESSION IS FIRST UPREGULATED IN SKELETAL
CC MUSCLE AT 12 DAYS PC (POST COITUM), THIS UPREGULATION IS EVIDENT
CC FIRST IN BODY WALL MUSCLE AND ONE DAY LATER IN LIMB MUSCLES. AT
CC 13.5 DAYS PC, A MOST PROMINENT EXPRESSION IS SEEN IN ALL SKELETAL
CC MUSCLES. AT THIS STAGE EXPRESSION IS SEEN IN ALL OTHER CELLS AND
CC TISSUES BUT AT LOWER LEVELS THAN IN SKELETAL MUSCLE.
CC -1- SIMILARITY: TO SKI PROTEIN OF SLOAN-KETTERING VIRUS AND TO SNO
CC ONCOGENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14173; AAA9669.1; -
DR MGD: MGI:98310; Ski.
DR InterPro: IPR003180; Transform_Ski.
DR Pfam: PF02437; Ski_Sno; 1.
KW Proto-oncogene; Nuclear protein; Repeat.
FT NON TER 348
SQ SEQUENCE 348 AA; 37869 MW; 4DD73478145D038C CRC64;

```

```

Query Match 27.1%; Score 52; DB 1; Length 348;
Best Local Similarity 35.3%; Pred. No. 6.9;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

```

```

OY 10 PVGPGEF-----KTLRVONLCGVVSG 33
   |||||
DB 82 PVLPGEFPPPSDRSTERCETVLEGGTISCFVVG 115

```

```

RESULT 8
SKI_AVIES STANDARD; PRT; 437 AA.
AC P17863;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transforming protein Ski.
GN v-Ski.
OS Avian erythroblastosis virus (strain Sloan-Kettering).
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11862;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=89384634; PubMed=2674685;
RA Stavezer E., Brodeur D., Brennan L.A.;
RT "The v-ski oncogene encodes a truncated set of c-ski coding exons
RT with limited sequence and structural relatedness to v-myc."

```

```

RL Mol. Cell. Biol. 9:4038-4045(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M28490; AAA9142.1; -.
DR PIR: A32574; TVFSK.
DR InterPro: IPR003380; Transform_Ski.
DR Pfam: PF02437; Ski_Sno; 1.
KW Oncogene.
SQ SEQUENCE 437 AA; 48867 MW; A29372928E0036F8 CRC64;

Query Match
Best Local Similarity 35.3%; Pred. No. 8.9;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGEF-----KTLRVONLCGYVSG 33
Db 46 PVMPGFPMPSDRSTERCETILBGETISCFVVG 79

RESULT 9
SKI_HUMAN STANDARD; PRT; 728 AA.
AC P1755;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ski oncogene (C-ski).
GN SKI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89345144; PubMed=2762147;
RA Nomura N., Sasamoto S., Ishii S., Matsui M., Ishizaki R.;
RT "Isolation of human cDNA clones of ski and the ski-related gene,
RT sno.";
RL Nucleic Acids Res. 17:5489-5500(1989).
CC -!- FUNCTION: MAY PLAY A ROLE IN TERMINAL DIFFERENTIATION OF SKELETAL
CC MUSCLE CELLS BUT NOT IN THE DETERMINATION OF CELLS TO THE MYOGENIC
CC LINEAGE.
CC -!- SUBUNIT: INTERACTS WITH SMAD2, 3 AND 4.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: TO SKI PROTEIN OF SLOAN-KETTERING VIRUS AND TO SNO
CC ONCOGENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15218; CAA33288.1; -.
DR PIR: S06053; TVHUSK.
DR TRANSFAC: T04643; -.
DR Genew: HGNC:10896; SKI.
DR MIM: 164780;
DR InterPro: IPR003380; Transform_Ski.
DR Pfam: PF02437; Ski_Sno; 1.
KW Proto-oncogene; Coiled coil; Nuclear protein; Repeat.
FT DOMAIN 536 710
SQ SEQUENCE 728 AA; 80004 MW; 9B78C44A0A28C2DA CRC64;

```

```

Query Match
Best Local Similarity 35.3%; Pred. No. 15;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGEF-----KTLRVONLCGYVSG 33
Db 84 PVLPGFPMPSDRSTERCETILBGETISCFVVG 117

RESULT 10
SKI_CHICK STANDARD; PRT; 750 AA.
AC P49140;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ski oncogene (C-ski).
GN SKI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89384635; PubMed=2779576;
RA Suttrave P., Hughes S.H.;
RT "Isolation and characterization of three distinct cDNAs for the
RT chicken c-ski gene.";
RL Mol. Cell. Biol. 9:4046-4051(1989).
CC -!- FUNCTION: MAY PLAY A ROLE IN TERMINAL DIFFERENTIATION OF SKELETAL
CC MUSCLE CELLS BUT NOT IN THE DETERMINATION OF CELLS TO THE MYOGENIC
CC LINEAGE.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: TO SKI PROTEIN OF SLOAN-KETTERING VIRUS AND TO SNO
CC ONCOGENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M28517; AAA8730.1; -.
DR InterPro: IPR003380; Transform_Ski.
DR Pfam: PF02437; Ski_Sno; 1.
KW Proto-oncogene; Coiled coil; Nuclear protein; Repeat.
FT DOMAIN 558 732
SQ SEQUENCE 750 AA; 84282 MW; 01E322EBE0C402B CRC64;

Query Match
Best Local Similarity 35.3%; Pred. No. 16;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGEF-----KTLRVONLCGYVSG 33
Db 67 PVMPGFPMPSDRSTERCETILBGETISCFVVG 100

RESULT 11
GLI_MOUSE STANDARD; PRT; 1111 AA.
AC P47806; OBOYK1;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein Gli1 (Glioma-associated oncogene homolog).
GN GLI OR GLI1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```



NCBI\_TaxID=10090;  
 (1)  
 SEQUENCE FROM N.A.  
 MEDLINE=98192509; PubMed=9524201;  
 Liu C.Z., Yang J.T., Yoon J.W., Walterhouse D., Iannaccone P.;  
 "Characterization of the promoter region and genomic organization of  
 GLI1, a member of the Sonic hedgehog-Patched signaling pathway.";  
 RL Gene 209:1-11(1998).  
 (2)  
 SEQUENCE FROM N.A.  
 STRAIN=ICR;  
 MEDLINE=99365280; PubMed=10433919;  
 Saeki H., Nishizaki Y., Hui C., Nakafuku M., Kondoh H.;  
 "Regulation of Gli2 and Gli3 activities by an amino-terminal  
 repression domain: implication of Gli2 and Gli3 as primary mediators  
 of Shh signaling.";  
 RL Development 126:3915-3924(1999).  
 (3)  
 SEQUENCE OF 272-837 FROM N.A.  
 STRAIN=CD-1; TISSUE=Embryo;  
 MEDLINE=99337281; PubMed=8364225;  
 Waterhouse D., Ahmed M., Slusarski D., Kalamara J., Boucher D.,  
 Holmgren R., Iannaccone P.;  
 "Gli1, a zinc finger transcription factor and oncogene, is expressed  
 during normal mouse development.";  
 RL Dev. Dyn. 196:91-102(1993).  
 -1- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING  
 NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT  
 AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL  
 NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE  
 TRANSDUCTION OF SHH SIGNAL.  
 -1- SUBCELLULAR LOCATION: Nuclear.  
 -1- TISSUE SPECIFICITY: UTERUS, BRAIN, TESTIS AND LIMB.  
 -1- DEVELOPMENTAL STAGE: IS DETECTED ON DAYS 10 THROUGH 18 OF  
 EMBRYONIC DEVELOPMENT. DURING GESTATION IT IS DETECTED IN MECKELS  
 PREGARTILAGE MESENCHYME, THE BASIS OCCIPITUS, RIB MESENCHYMAL  
 CONDENSATIONS, PRIMORDIAL VERTEBRAL BODIES, DIGITAL MESENCHYMAL  
 CONDENSATIONS IN FOREFOOT AND HINDFOOT PLATES, THE EPENDYMAL LAYER  
 OF THE SPINAL CORD, AND THE MESODERM OF THE GASTROINTESTINAL  
 TRACT. EXPRESSION PERSISTS THROUGHOUT GESTATION IN DEVELOPING BONE  
 AND CARTILAGE OF THE EXTREMITIES, THE RIBS, AND THE VERTEBRAL  
 BODIES AS WELL AS THE GASTROINTESTINAL TRACT MESODERM.  
 -1- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER  
 PROTEINS.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL, AF026305; AAC09169.1; -;  
 EMBL, AB025922; BAA85004.1; -;  
 DR HSSP, P08151; ZGLI.  
 DR MGD, MGI:95727; Gli.  
 DR InterPro, IPR000822; Znf C2H2.  
 DR Pfam, PF00096; Zf-C2H2; 5.  
 DR SMART, SM00355; ZNF\_C2H2; 5.  
 DR PROSITE, PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
 DR PROSITE, PS50157; ZINC\_FINGER\_C2H2\_2; 5.  
 KW Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;  
 KW Nuclear protein.  
 FT DOMAIN 228 390 ZINC\_FINGERS.  
 FT ZN\_FING 228 263 C2H2-TYPE.  
 FT ZN\_FING 271 298 C2H2-TYPE.  
 FT ZN\_FING 304 328 C2H2-TYPE.  
 FT ZN\_FING 334 359 C2H2-TYPE.  
 FT ZN\_FING 365 390 C2H2-TYPE.  
 FT DOMAIN 1042 1059 ASP/GLU-RICH (ACIDIC).  
 FT CONFLICT 154 154 V -> L (IN REF. 1).  
 FT CONFLICT 170 170 H -> Y (IN REF. 1).

FT CONFLICT 173 173 A -> S (IN REF. 1).  
 FT CONFLICT 179 179 T -> I (IN REF. 1).  
 FT CONFLICT 194 194 P -> R (IN REF. 1).  
 FT CONFLICT 210 210 I -> T (IN REF. 1).  
 FT CONFLICT 271 271 F -> S (IN REF. 1).  
 FT CONFLICT 475 475 MISSING (IN REF. 1 AND 3).  
 FT CONFLICT 567 569 FPP -> LPT (IN REF. 1 AND 3).  
 FT CONFLICT 707 707 E -> D (IN REF. 1 AND 3).  
 FT CONFLICT 777 777 MISSING (IN REF. 1 AND 3).  
 FT CONFLICT 837 837 H -> P (IN REF. 1 AND 3).  
 FT CONFLICT 864 864 G -> V (IN REF. 1).  
 FT CONFLICT 919 920 GU -> RA (IN REF. 1).  
 FT CONFLICT 936 936 S -> Y (IN REF. 1).  
 FT CONFLICT 950 952 AAA -> RR (IN REF. 1).  
 FT CONFLICT 967 967 G -> R (IN REF. 1).  
 FT CONFLICT 990 990 P -> A (IN REF. 1).  
 FT CONFLICT 1029 1029 A -> P (IN REF. 1).  
 FT CONFLICT 1062 1063 OG -> R (IN REF. 1).  
 SQ SEQUENCE 1111 AA; 118555 MW; 0DBAB163CAFA27289 CRC64;  
 Query Match 27.1%; Score 52; DB 1; Length 1111;  
 Best Local Similarity 60.0%; Pred. No. 24;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 YEHPYGLPYGPGP 15  
 Db 744 YEARGPGSLPFGPGP 758  
 RESULT 12  
 ID CP5G CANTR STANDARD; PRT; 507 AA.  
 AC P30609;  
 DT 01-APR-1993 (Rel. 25. Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytochrome P450 52A7 (BC 1.14.14.-) (CYPL1A7) (Alkane-inducible P450-  
 ALK4).  
 DE CYP52A7.  
 GN Candida tropicalis (Yeast).  
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=54482;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 750;  
 RX MEDLINE=93090274; PubMed=1457045;  
 RA Seghezzi W., Meli C., Ruffiner R., Kuenzi R., Sanglard D.,  
 R Flechter A.;  
 RT "Identification and characterization of additional members of the  
 cytochrome P450 multigene family CYP52 of Candida tropicalis.";  
 RL DNA Cell Biol. 11:767-780(1992).  
 CC -1- FUNCTION: TOGETHER WITH AN NADPH CYTOCHROME P450 THE ENZYME SYSTEM  
 CATALYZES THE TERMINAL HYDROXYLATION AS THE FIRST STEP IN THE  
 ASSIMILATION OF ALKANES AND FATTY ACIDS. PREFERENTIALLY  
 CC HYDROXYLATES LAURIC ACID.  
 CC -1- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal  
 position.  
 CC -1- INDUCTION: BY VARIOUS ALKANES.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL, Z13011; CAA78355.1; -;  
 DR PIR, S22973; S22973.  
 DR HSSP, P14779; 10PZ.  
 DR InterPro, IPR001128; Cytochrome\_P450.

DR Pfam: PF00067; P450; 1.  
 DR PRINTS: PR00385; P450.  
 DR PROSITE: PS00086; CYTOCHROME P450; 1.  
 KW Electron transport; Oxidoreductase; Monooxygenase; Heme;  
 FT TRANSMEM 6 POTENTIAL.  
 FT BINDING 456 456 HEME (BY SIMILARITY).  
 SQ SEQUENCE 507 AA; 58671 MW; 34833DE3D1E5C87 CRC64;

Query Match  
 Best Local Similarity 40.7%; Score 51.5; DB 1; Length 507;  
 Matches 11; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

QY 8 TLPGGPEPFTLRVONLCGYV-VSG 33  
 DB 388 TLPGGPEPFTLRVONLCGYV-VSG 414

RESULT 13  
 YD29 SCHPO STANDARD; PRT; 402 AA.  
 ID YD29 SCHPO  
 AC 013716;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C14C4.09 in chromosome 1.  
 GN SPAC14C4.09.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RC MEDLINE=21849401; PubMed=1859360;  
 RX Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,  
 RA Welfens I., Vansireels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wandut R., Fumelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucass M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Banito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe".  
 RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z89596; CAB11202.1;  
 DR InterPro: IPR005197; Glyco\_hydro\_71.

DR Pfam: PF03659; Glyco\_hydro\_71; 1.  
 KW Hypothetical protein.  
 KW SEQUENCE 402 AA; 44515 MW; DCB13CDB85E070B0 CRC64;  
 SQ SEQUENCE 1053 AA; 118292 MW; D6C933798567820A CRC64;

Query Match  
 Best Local Similarity 43.5%; Score 51; DB 1; Length 402;  
 Matches 10; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 7 GTLPVGPPEPFTLRVONLCGY 29  
 DB 374 GTLPVGPPEPFTLRVONLCGY 396

RESULT 14  
 SPS SOLTU STANDARD; PRT; 1053 AA.  
 ID SPS SOLTU  
 AC 043845;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Sucrose-phosphate synthase (EC 2.4.1.14) (UDP-glucose-fructose-  
 DE phosphate glucosyltransferase).  
 GN SPS.  
 OS Solanum tuberosum (Potato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; easterids I; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Deslee; TISSUE=Leaf;  
 RX MEDLINE=95201832; PubMed=7894514;  
 RA Zrenner R., Salanoubat M., Willmitzer L., Somewald U.,  
 RT "Evidence of the crucial role of sucrose synthase for sink strength  
 RT using transgenic potato plants (Solanum tuberosum L.).".  
 RL Plant J. 7:97-107(1995).  
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN  
 CC THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND  
 CC THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF  
 CC PHOTOSYNTHETATES OUT OF THE LEAF.  
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +  
 CC sucrose 6-phosphate.  
 CC -1- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND  
 CC MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.  
 CC -1- PATHWAY: Sucrose synthesis.  
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER (BY SIMILARITY).  
 CC -1- PFM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR  
 CC ENZYME FUNCTION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X73477; CAAS1872.1;  
 DR InterPro: IPR001296; Glycos\_transf\_1.  
 DR Pfam: PF00534; Glycos\_transf\_1; 1.  
 KW Transferase; Glycosyltransferase; Phosphorylation.  
 SQ SEQUENCE 1053 AA; 118292 MW; D6C933798567820A CRC64;

Query Match  
 Best Local Similarity 35.5%; Score 51; DB 1; Length 1053;  
 Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEHPHYGTLPGGPEPFTLRVONLCGYV 31  
 DB 911 FKVCKPGTVP--PSKELRKVRIOALRCHAV 939

ID	NCBI_MOUSE	STANDARD;	PRT; 1064 AA.
AC	Q60591; Q60985; Q60984;		
AD	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Nuclear factor of activated T-cells, cytoplasmic 2 (T cell)		
DE	transcription factor NFAT1 (NFAT pre-existing subunit) (NF-ATp).		
GN	NFATC2 OR NFAT1 OR NFATP.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM A).		
RX	MEDLINE=94053710; PubMed=8235597;		
RA	McCaffrey P.G., Luo C., Kerpolla T.K., Jain J., Badalian T.M.,		
RA	Ho A.M., Burgeon E., Lane W.S., Lambert J.N., Curran T., Verdine G.L.,		
RA	Rao A., Hogan P.G.;		
RT	"Isolation of the cyclosporin-sensitive T cell transcription factor		
RT	NFATP.";		
RL	Science 262:750-754(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORMS A; B AND C).		
RX	MEDLINE=96251346; PubMed=8668213;		
RA	Luo C., Burgeon E., Carew J.A., McCaffrey P.G., Badalian T.M.,		
RA	Lane W.S., Hogan P.G., Rao A.;		
RT	"Recombinant NFAT1 (NFATp) is regulated by calcineurin in T cells and		
RT	mediates transcription of several cytokine genes.";		
RL	Mol. Cell. Biol. 16:3955-3966(1996).		
RN	[3]		
RP	MUTAGENESIS OF ARG-423; HIS-425; TYR-426; THR-428 AND GLU-429.		
RX	MEDLINE=95181386; PubMed=7876165;		
RA	Jain J., Burgeon E., Badalian T.M., Hogan P.G., Rao A.;		
RT	"A similar DNA-binding motif in NFAT family proteins and the Rel		
RT	homology region.";		
RL	J. Biol. Chem. 270:4138-4145(1995).		
RN	[4]		
RP	MUTAGENESIS OF ARG-112; GLU-114 AND THR-116.		
RX	MEDLINE=98325377; PubMed=9660947;		
RA	Aramburu J., Garcia-Cozar F., Raghavan A., Okamura H., Rao A.,		
RA	Hogan P.G.;		
RT	"selective inhibition of NFAT activation by a peptide spanning the		
RT	calcineurin targeting site of NFAT.";		
RL	Mol. Cell 1:627-637(1998).		
RN	[5]		
RP	REVIEW		
RX	MEDLINE=99189746; PubMed=10089876;		
RA	Ciabrey G.R.;		
RT	"Genetic signals and specific outcomes: signaling through Ca2+,		
RT	calcineurin, and NF-AT.";		
RL	Cell 96:611-614(1999).		
CC	-1- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE		
CC	GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2, IL-3,		
CC	IL-4, TNF-ALPHA OR GM-CSF.		
CC	-1- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX		
CC	THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING		
CC	CYTOPLASMIC COMPONENT NFATC AND AN INDUCIBLE NUCLEAR COMPONENT		
CC	NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE		
CC	ACTIVATING PROTEIN-1 FAMILY, MAP, GATA4 AND C/EB/300 CAN ALSO BIND		
CC	THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS.		
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND		
CC	NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-		
CC	MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT		
CC	TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED		
CC	AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF		
CC	NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.		
CC	-1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A (SHOWN HERE), B AND C; ARE		
CC	PRODUCED BY ALTERNATIVE SPLICING.		
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, SPLEEN, HEART, TESTIS,		
CC	BRAIN, PLACENTA, MUSCLE AND PANCREAS.		
CC	-1- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND		

```

CC COOPERATIVE INTERACTIONS WITH A21 FACTORS (BY SIMILARITY) .
CC -1- PPM: PHOSPHORYLATED BY NEATC-KINASE, DEPHOSPHORYLATED BY
CC CALCINEURIN (BY SIMILARITY)
CC -1- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL; U02079; AAC52929.1; -
CC EMBL; U36575; AAC52930.1; -
CC EMBL; U36576; AAC52931.1; -
CC MGD; MGI:102463; NFATc2
CC InterPro; IPRO00451; NF_Rel_dor_fam.
CC InterPro; IPRO02909; IPT_TIG.
CC Pfam; PF00554; RHD; 1.
CC Pfam; PF01833; TIG; 1.
CC SMART; SMO0429; IPT; 1.
CC PROSITE; PS01204; REL_1; FALSE_NEG.
CC PROSITE; PS50254; REL_2; 1.
CC Transcription regulation; Activator; Nuclear protein; DNA-binding;
CC Alternative splicing; Phosphorylation; Repeat.
CC DOMAIN 111 116
CC DOMAIN 119 201
CC DOMAIN 186 292
CC REPEAT 186 202
CC REPEAT 215 231
CC REPEAT 274 290
CC DOMAIN 423 430
CC MOD_RS 110 110
CC VARSPIC 910 1064
CC
CC PHOSPHORYLATION (BY SIMILARITY).
CC AATESVWGERYIERFERFKKTLVQPGILPSFLIGSLASG
CC PROTPSEKRIEDVPLSCQIAMCCOHPGTCPLVGGPL
CC AVEWMEQGLGRLEIPWAPDSAGSLHEVDSVGLAGVGMV
CC LITLTHHMSDONOTPSPHMORHKEVASPGMI -> ELIDT
CC HLSTQHL (IN ISOFORM B).
CC AATESVWGERYIERFERFKKTLVQPGILPSFLIGSLASG
CC PROTPSEKRIEDVPLSCQIAMCCOHPGTCPLVGGPL
CC AVEWMEQGLGRLEIPWAPDSAGSLHEVDSVGLAGVGMV
CC LITLTHHMSDONOTPSPHMORHKEVASPGMI -> ELIDT
CC HLSTQHL (IN ISOFORM C).
CC RKEESGPSNRQT (IN ISOFORM C).
CC E->A: LOWERS DEPHOSPHORYLATION.
CC E->A: LOWERS DEPHOSPHORYLATION.
CC T->A: NO DEPHOSPHORYLATION.
CC R->A: DECREASE IN-BINDING TO DNA.
CC H->A: NO CHANGE IN BINDING TO DNA.
CC Y->A: DECREASE IN BINDING TO DNA.
CC T->A: NO CHANGE IN BINDING TO DNA.
CC T->C: NO CHANGE IN BINDING TO DNA AND
CC CONFERES DNA-BINDING SENSITIVITY TO
CC SUPHYDRYL MODIFICATIONS.
CC E->A: DECREASE IN BINDING TO DNA.
CC B->A: DECREASE IN BINDING TO DNA.
CC
CC MUTAGEN 429 429
CC SEQUENCE 1064 AA; 115028 MW; C15C166D1D3CA6F0 CRC64;
CC
CC Query Match 26.6%; Score 51; DB 1; Length 1064;
CC Best Local Similarity 75.0%; Pred. No. 32;
CC Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
CC
CC 4 HPVGTLPVGPFP 15
CC ||| ||| |||
CC ||| ||| |||
CC 979 HPLGTCEVLPGP 990

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:02 ; Search time 3.59353 Seconds  
(without alignments)  
2006.842 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192

Sequence: 1 YEHHPYGTLPVGPGBEFKTLRVONLGCYVSGLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	30.7	823	4 Q8WY18	Q8WY18 homo sapien
2	58.5	30.5	2275	12 Q8Q541	Q8Q541 chimpanzee
3	58	30.2	695	13 Q9Y102	Q9Y102 brachydanio
4	56	29.2	94	15 Q9Y101	Q9Y101 human immun
5	55	28.6	1054	10 Q9SNY7	Q9SNY7 nicotiana t
6	54.5	28.4	1173	17 Q8ZX19	Q8ZX19 pyrobaculum
7	54	28.1	699	5 O15816	O15816 dictyosteli
8	53	27.6	94	15 Q9Y121	Q9Y121 human immun
9	53	27.6	714	13 Q9P031	Q9P031 oreochromis
10	53	27.6	727	13 Q9Y103	Q9Y103 brachydanio
11	52	27.1	94	15 Q9Y125	Q9Y125 human immun
12	52	27.1	359	13 Q8UMC1	Q8UMC1 gallus gall
13	52	27.1	461	17 Q979H6	Q979H6 thermoplasma
14	52	27.1	725	11 Q8V115	Q8V115 mus musculu
15	52	27.1	730	6 Q9TUG2	Q9TUG2 equus caball

17	51.5	26.8	390	5 P91571	P91571 caenorhabdi
18	51	26.6	94	15 Q9Y134	Q9Y134 human immun
19	51	26.6	312	12 Q8UZD8	Q8UZD8 cercopithec
20	51	26.6	366	7 Q8SNF2	Q8SNF2 sigmodon hi
21	51	26.6	384	13 Q73714	Q73714 brachydanio
22	51	26.6	741	11 Q9CY19	Q9CY19 mus musculu
23	51	26.6	741	11 Q9ROE1	Q9ROE1 mus musculu
24	50.5	26.3	300	2 Q68031	Q68031 rhodospacer
25	50.5	26.3	317	16 Q9PD79	Q9PD79 xylella fas
26	50	26.0	93	15 Q9Y153	Q9Y153 human immun
27	50	26.0	94	15 Q9Y155	Q9Y155 human immun
28	50	26.0	94	15 Q9Y146	Q9Y146 human immun
29	50	26.0	94	15 Q9Y145	Q9Y145 human immun
30	50	26.0	94	15 Q9Y140	Q9Y140 human immun
31	50	26.0	94	15 Q9Y139	Q9Y139 human immun
32	50	26.0	94	15 Q9Y137	Q9Y137 human immun
33	50	26.0	94	15 Q9Y131	Q9Y131 human immun
34	50	26.0	94	15 Q9Y126	Q9Y126 human immun
35	50	26.0	94	15 Q9Y153	Q9Y153 human immun
36	50	26.0	94	15 Q9Y155	Q9Y155 human immun
37	50	26.0	142	2 Q9AH33	Q9AH33 pseudomonas
38	50	26.0	142	2 Q9RIH7	Q9RIH7 pseudomonas
39	50	26.0	149	10 Q9FSN5	Q9FSN5 oryza sativ
40	50	26.0	191	17 Q8TVD9	Q8TVD9 methanopyru
41	49.5	25.8	94	15 Q9Y141	Q9Y141 human immun
42	49.5	25.8	348	17 Q97V15	Q97V15 sulfolobus
43	49.5	25.8	361	6 Q95J14	Q95J14 macaca fasc
44	49.5	25.8	398	5 Q8SVW6	Q8SVW6 encephalito
45	49.5	25.8	585	6 Q95LZ0	Q95LZ0 macaca fasc

## ALIGNMENTS

RESULT 1	Q8WY18	PREDIMINARY;	PRT;	823 AA.
AC	Q8WY18;			
DT	01-MAR-2002 (TREMBlrel. 20, Created)			
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)			
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)			
DE	MSTP018.			
GN	MST018.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=AOA;			
RA	Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,			
RA	Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.D.,			
RA	Gao R.L., Qiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,			
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF11799; AAL39001.1; -			
DR	InterPro; IPR000413; Integrin_alpha.			
DR	Pfam; PF01839; FG-GAP; 4.			
DR	PRINTS; PRO1185; INTEGRIN.			
DR	SMART; SM00191; Int_alpha; 4.			
DR	SEQUENCE 823 AA; 92672 MW; DE4E78079DCD4925 CRC64;			
Query Match	30.7%; Score 59; DB 4; Length 823;			
Best Local Similarity	41.7%; Pred. No. 7.4;			
Matches	10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;			
Qy	12 GPGEFKTLRVONLGCYVSGLI 35			
DB	616 GIGPPEFSCIRIQLGLFPIHGM 639			
RESULT 2	Q8Q541			

ID Q8Q541 PRELIMINARY; PRT; 2275 AA.  
 AC Q8Q541;  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Tegument protein UL48.  
 OS Chimpanzee cytomegalovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OC NCBI\_TaxID=188763;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Davison A.J., Akter P., Dolan A., Wright K.M., Addison C.,  
 RA Alencor D.J., Hayward G.S., McGeoch D.J.;  
 RA "The human cytomegalovirus genome revisited."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF480884; AF480697.1;  
 SQ SEQUENCE 2275 AA; 255990 MW; 45BBA419CA576BCD CRC64;

Query Match 30.5%; Score 58.5; DB 12; Length 2275;  
 Best Local Similarity 34.0%; Pred. No. 28;  
 Matches 16; Conservative 7; Mismatches 7; Indels 17; Gaps 3;

OY 1 YEVHPY-----GTLPVGP--PEFKTLRVONTGCVV 31  
 DB 1680 YDLKPYFSQQQQQRAQDGSVPLGPAPPEANTLNLK-LFCYV 1725

## RESULT 3

O9Y102 PRELIMINARY; PRT; 695 AA.  
 ID Q9Y102;  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Nuclear oncoprotein.  
 GN SKIB OR SKIB.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OC NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20368164; PubMed=10906458;  
 RA Kaufman C.D., Martinez-Rodriguez G., Hackelt P.B. Jr;  
 RT "Ecologic expression of c-ski disrupts gastrulation and neural  
 RT patterning in zebrafish."  
 RL Mech. Dev. 95:147-162(2000).  
 DR EMBL; AF060118; AAC64707.1;  
 DR ZFIN; ZDB-GENE-990715-10; skib.  
 DR InterPro: IPR003380; Transform\_Ski.  
 DR Pfam; PF02437; Ski\_Sno; 1.  
 SQ SEQUENCE 695 AA; 79135 MW; 9BD424C32F61C4ED CRC64;

Query Match 30.2%; Score 58; DB 13; Length 695;  
 Best Local Similarity 35.9%; Pred. No. 8.5;  
 Matches 14; Conservative 4; Mismatches 11; Indels 10; Gaps 1;

OY 5 PYGTLPVGPFP-----KTLRVONTGCVVSG 33  
 DB 58 PAQTPVMPGPFIPSDRSTERCETVLERETISCFVVG 96

## RESULT 4

O9YTS1 PRELIMINARY; PRT; 94 AA.  
 ID Q9YTS1;  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Envelope-glycoprotein (Fragment).  
 GN ENV.

OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S4V2-6;  
 RX MEDLINE=98445411; PubMed=9770526;  
 RA Markham R.B., Wang W.C., Weisstein A.E., Wang Z., Munoz A.,  
 RA Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H.,  
 RA Xu X.F.;  
 RT "Patterns of HIV-1 evolution in individuals with differing rates of  
 RT CD4 T cell decline."  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).  
 DR EMBL; AF089156; AAC78866.1;  
 DR InterPro: IPR000777; GP120.  
 DR Pfam; PF00516; GP120.1.  
 KM AIDS; Coat protein; Glycoprotein.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 94 AA; 10841 MW; 27E1BE1608E890C4 CRC64;

Query Match 29.2%; Score 56; DB 15; Length 94;  
 Best Local Similarity 47.4%; Pred. No. 1.7;  
 Matches 9; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 8 TLVPVGPGEFKTLRVONTL 26  
 DB 38 SIPIGPGRAFTTGRINKI 56

## RESULT 5

O9SNY7 PRELIMINARY; PRT; 1054 AA.  
 ID O9SNY7;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Sucrose-6-phosphate synthase (EC 2.4.1.14).  
 GN SPS.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 OC NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. SAMSUN NN;  
 RA Borker F.;  
 RT "Cloning and Characterization of a Sucrose-6-phosphate Synthase from  
 RT Nicotiana tabacum."  
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF194032; AAF06792.1;  
 DR InterPro: IPR001296; Glycosyltransf\_1.  
 DR Pfam; PF00534; Glycosyltransf\_1.  
 DR KMW Glycosyltransferase; Transferrase.  
 SQ SEQUENCE 1054 AA; 118692 MW; 2C593AF27649A26D CRC64;

Query Match 28.6%; Score 55; DB 10; Length 1054;  
 Best Local Similarity 35.5%; Pred. No. 37;  
 Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

OY 1 YEVHPYGLPVGPGEFKTLRVONTGCVV 31  
 DB 912 FKVKRLGTV--PAKELRLKMLRIQALCHAV 940

## RESULT 6

O8ZXI9 PRELIMINARY; PRT; 1173 AA.  
 ID O8ZXI9;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Molybdopter in oxidoreductase, molybdopter in binding subunit.  
GN PAE1265.  
OS Pyrobaculum aerophilum.  
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
CC Thermoproteaceae; Pyrobaculum.  
OX NCBI\_TaxId=13773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
RX PubMed=11792869;  
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
Miller J.H.;  
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
aerophilum.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
DR EMBL; AB009811; AAL6359.1; -.  
DR InterPro; IPR000566; Lipocin\_cytfABP.  
DR InterPro; IPR001467; Prok\_Mboxred.  
DR Pfam; PF00384; molybdopterin; 1.  
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 1173 AA; 132457 MW; 790FCC747D512BC CRC64;  
  
QY Query Match 28.4%; Score 54.5; DB 17; Length 1173;  
Best Local Similarity 41.4%; Pred. No. 50;  
Matches 12; Conservative 4; Mismatches 10; Indels 3; Gaps 1;  
DB 10 PVGPGPEFKTLRV---QNLGCVVSGLI 35  
38 PYGPDQFGTNRVYVSSCLGCVRCGIV 66  
  
RESULT 7  
ID 015816 PRELIMINARY; PRT; 699 AA.  
AC 015816;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Myb2 protein.  
GN MYB2.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxId=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX3; PubMed=9620859;  
RX MEDLINE=96283920; PubMed=9620859;  
RA Otsuka H., Van Haastert P.J.M.;  
RT "A novel Myb homologue initiates Dictyostelium development by  
induction of adenyl cyclase expression.";  
RL Genes Dev. 12:1738-1748(1998).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 3 MYB-LIKE DOMAINS.  
DR EMBL; AJ002383; CAA05357.1; -.  
DR HSSP; P06876; IMBG.  
DR InterPro; IPR001005; Myb DNA-binding.  
DR Pfam; PF00249; myb DNA-binding; 3.  
DR SMART; SM00395; SANT; 3.  
DR PROSITE; PS50090; MYB\_3; 3.  
KW DNA-binding; Nuclear protein.  
SQ SEQUENCE 699 AA; 80381 MW; EA989DC77953F265 CRC64;  
  
QY Query Match 28.1%; Score 54; DB 5; Length 699;  
Best Local Similarity 37.5%; Pred. No. 32;  
Matches 12; Conservative 3; Mismatches 17; Indels 0; Gaps 0;  
DB 4 HPYGTLPVGPGEFKTLRVONLIGCVVSGLI 35  
55 HHHONIPRPDPIFSPQILINKKPGCYGTSNOI 86

RESULT 8  
ID 09YT21 PRELIMINARY; PRT; 94 AA.  
AC 09YT21;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxId=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S4V4-10;  
RX MEDLINE=98445411; PubMed=9770526;  
RA Markham R.B., Wang W.C., Weinstein A.E., Wang Z., Munoz A.,  
Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H.,  
Yu X.F.;  
RT "Patterns of HIV-1 evolution in individuals with differing rates of  
CD4 T cell decline.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).  
DR EMBL; AF089191; AAC78901.1; -.  
DR InterPro; IPR000777; GP120.  
DR Pfam; PF00516; GP120; 1.  
DR AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1 1  
FT NON\_TER 94 94  
SQ SEQUENCE 94 AA; 10797 MW; 22C21B4D098FA672 CRC64;  
  
QY Query Match 27.6%; Score 53; DB 15; Length 94;  
Best Local Similarity 43.5%; Pred. No. 46;  
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
DB 4 HPYGTLPVGPGEFKTLRVONL 26  
34 HTVRKIPIGPSSFYTGVRGDI 56  
  
RESULT 9  
ID 09PU31 PRELIMINARY; PRT; 714 AA.  
AC 09PU31;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE C-ski protein.  
GN C-SKI.  
OS Oreochromis aureus (Israeli tilapia).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorphi; Acanthopterygii; Perciformes; Labroidae;  
OC Cichlidae; Oreochromis.  
OX NCBI\_TaxId=47969;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Huang C.J., Lin J.Y., Tsai H.J.;  
RT "Two distinct c-ski proto-oncogene cDNAs of fish, tilapia (Oreochromis  
aurea).";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ012012; CAB58126.1; -.  
DR InterPro; IPR003380; Transform\_Ski.  
DR Pfam; PF02437; Ski\_Sno; 1.  
SQ SEQUENCE 714 AA; 81076 MW; F2F8AB6607CCC72F CRC64;  
  
QY Query Match 27.6%; Score 53; DB 13; Length 714;  
Best Local Similarity 35.3%; Pred. No. 46;  
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;  
DB 10 PVGPGPEF-----KTLRVONLIGCVVSG 33  
63 FVMPGPLFVPSDRSTERCETVLEGETISCFVVG 96

## RESULT 10

Q9Y103 PRELIMINARY; PRT; 727 AA.  
 AC Q9Y103; 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, last sequence update)  
 DE 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
 GN Nuclear oncoprotein.  
 OS SKIA OR SKIA.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OC NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20368164; PubMed=10906458;  
 RA Kautman C.D., Martinez-Rodriguez G., Hackett P.B. Jr.;  
 RT "Ectopic expression of c-ski disrupts gastrulation and neural  
 RT patterning in zebrafish";  
 RL Mech. Dev. 95:147-162(2000).  
 DR EMBL; AF060117; AAC64706.1; -;  
 DR ZFIN; ZDB-GENE-990715-9; skia.  
 DR InterPro; IPR003380; Transform\_Skt.  
 DR Pfam; PF02437; Ski\_Sno; 1.  
 SQ SEQUENCE 727 AA; 80973 MW; EDEED8EDCCFALA70 CRC64;

Query Match 27.1%; Score 53; DB 13; Length 727;  
 Best Local Similarity 35.3%; Pred. No. 47;  
 Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

OY 10 PVGPGPEF-----KTLRVONLGCYVSG 33  
 DB 67 PVMGPELIPSDRSTERCIVLEGRTISCFVGG 100

## RESULT 11

Q9YT25 PRELIMINARY; PRT; 94 AA.  
 AC Q9YT25; 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE Envelope glycoprotein (Fragment).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S4V4-4;  
 RX MEDLINE=98445411; PubMed=9770526;  
 RA Marham R.B., Wang W.C., Weisstein A.E., Wang Z., Munoz A.,  
 RA Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H.,  
 RA Yu X.F.;  
 RT "Patterns of HIV-1 evolution in individuals with differing rates of  
 RT CD4 T cell decline";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).  
 DR EMBL; AF089185; AAC78895.1; -;  
 DR InterPro; IPR000777; GP120.  
 DR Pfam; PF00516; GP120; 1.  
 DR AIDS; Coat protein; Glycoprotein.  
 DR InterPro; IPR001503; GT\_10.  
 DR Pfam; PF00852; Glyco\_rtmef\_10; 1.  
 FT NON\_TER  
 SQ SEQUENCE 94 AA; 10750 MW; DF7420C1A6179C68 CRC64;

Query Match 27.1%; Score 52; DB 15; Length 94;  
 Best Local Similarity 50.0%; Pred. No. 6.4;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 9 LPVGPGEFFKTLRVONL 26

DB 39 IPGPGRAFYTTGRIGNI 56

## RESULT 12

Q9YT23 PRELIMINARY; PRT; 94 AA.  
 AC Q9YT23; 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE Envelope glycoprotein (Fragment).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S4V4-7;  
 RX MEDLINE=98445411; PubMed=9770526;  
 RA Marham R.B., Wang W.C., Weisstein A.E., Wang Z., Munoz A.,  
 RA Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H.,  
 RA Yu X.F.;  
 RT "Patterns of HIV-1 evolution in individuals with differing rates of  
 RT CD4 T cell decline";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).  
 DR EMBL; AF089188; AAC78898.1; -;  
 DR InterPro; IPR000777; GP120.  
 DR Pfam; PF00516; GP120; 1.  
 DR AIDS; Coat protein; Glycoprotein.  
 DR NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 94 AA; 10781 MW; 249420C1B250B370 CRC64;

Query Match 27.1%; Score 52; DB 15; Length 94;  
 Best Local Similarity 50.0%; Pred. No. 6.4;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 9 LPVGPGEFFKTLRVONL 26  
 DB 39 IPGPGRAFYTTGRIGNI 56

## RESULT 13

Q8UWCI PRELIMINARY; PRT; 359 AA.  
 AC Q8UWCI; 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
 DE CFU9.  
 GN Gallus gallus (Chicken).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Kaneko M., Nishihara S., Kitano T., Narimatsu H., Saitou N.,  
 RA "The evolutionary history of glycosyltransferase genes";  
 RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB035906; BAB82489.1; -;  
 DR InterPro; IPR001503; GT\_10.  
 DR Pfam; PF00852; Glyco\_rtmef\_10; 1.  
 SQ SEQUENCE 359 AA; 42077 MW; 2A19DACD1F49B6C8 CRC64;

Query Match 27.1%; Score 52; DB 13; Length 359;  
 Best Local Similarity 42.9%; Pred. No. 30;  
 Matches 12; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

OY 5 PYGTLPVGPGEFFKTLRVONLGCYVVS 32



Db 167 PYGSMIVGTGATFTEVPSKENILVCWVS 194

## RESULT 14

0979H6 PRELIMINARY; PRT; 461 AA.

AC 0979H6; 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Shikimate transporter.  
 GN TV1185 OR TVG1213986.  
 OS Thermoplasma volcanium.  
 OC Archaeae; Euryarchaeota; Thermoplasmatia; Thermoplasmatales;  
 OC Thermoplasmataceae; Thermoplasma.  
 OX NCBI\_TaxID=50339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GSSI / DSM 4299 / JCM 9571;  
 RX MEDLINE=20570466; PubMed=11121031;  
 RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,  
 RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,  
 RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.,  
 RT "Archaeal adaptation to higher temperatures revealed by genomic  
 RT sequence of Thermoplasma volcanium."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).  
 DR EMBL; AP000995; BAB60327.1; -  
 DR InterPro; IPR003562; sub transporter.  
 DR Pfam; PF00083; sugar tr; 1  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 461 AA; 50273 MW; F4E80954FD218F4 CRC64;

Query Match 27.1%; Score 52; DB 17; Length 461;

Best Local Similarity 36.7%; Pred. No. 39;  
 Matches 11; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 5 PYGTLPGVGPGEFKTTLRVQNLGCVVSG 34  
 DB 371 PWGIVPVYLSERFKAIVRASGVGFGVSSGI 400

## RESULT 15

08VIL5 PRELIMINARY; PRT; 725 AA.

AC 08VIL5; 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Ski proto-oncogene.  
 GN SKI.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV.  
 RA Chen Y., Berk M., Chen H., Stavnezer E., Colmenares C.,  
 RT "Mouse Ski proto-oncogene cDNA."  
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF435852; AAL30825.1; -  
 DR InterPro; IPR003380; Transform\_Ski.  
 DR Pfam; PF02437; Ski\_Smc; 1.  
 SQ SEQUENCE 725 AA; 80119 MW; 1BFD05C38519505C CRC64;

Query Match 27.1%; Score 52; DB 11; Length 725;

Best Local Similarity 35.3%; Pred. No. 66;  
 Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

OY 10 PYGPGPER-----KTLRVQNLGCVVSG 33  
 DB 82 PVLPGPFPMPSDRSTERCEIVLEGETISCFVVG 115

Search completed: July 16, 2003, 07:55:28  
 Job time : 6.59353 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:22 ; Search time 3.88889 Seconds

(without alignments)  
1199.256 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192

Sequence: 1 YEHAPYGTLPVGPGEFXTTLRVNLCGYVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID2/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	100.0	1152	22	AAB64657
2	192	100.0	1152	22	AAB64658
3	192	100.0	1167	21	AAV32242
4	192	100.0	1167	22	AAB64584
5	142	74.0	1132	21	AAV32243
6	69	35.9	1188	22	AAU14231
7	69	35.9	1188	22	AAU14467
8	69	35.9	1188	22	AAB50085
9	69	35.9	1188	23	AAU10551
10	59	30.7	1034	21	AAB25590

11	59	30.7	1188	22	AAB30929
12	59	30.7	1189	21	AAB25562
13	59	30.7	1189	22	ABG12949
14	56	29.2	347	12	AAU11069
15	56	29.2	545	23	ABB72268
16	56	29.2	688	23	ABB72300
17	56	29.2	696	23	ABB72289
18	56	29.2	1188	22	AAB50087
19	56	29.2	1188	23	AAU10552
20	55	28.6	210	22	AAU62597
21	54	28.1	257	22	ABG20619
22	54	28.1	391	22	ABG20618
23	53.5	27.9	349	22	ABG14393
24	52.5	27.3	267	22	ABG28949
25	52	27.1	346	12	AAU11068
26	52	27.1	728	22	AAU40251
27	52	27.1	728	22	AAE05155
28	52	27.1	750	12	AAU14048
29	52	27.1	750	12	AAU15160
30	52	27.1	750	12	AAU10458
31	52	27.1	750	22	AAE01022
32	52	27.1	750	22	AAE01025
33	52	27.1	752	22	AAU42037
34	52	27.1	1111	23	AAU17108
35	51	26.6	106	23	ABU05801
36	51	26.6	741	23	AAU74438
37	51	26.6	846	15	AAU47474
38	51	26.6	890	15	AAU60252
39	51	26.6	890	22	AAU78052
40	50.5	26.3	462	22	AAU30882
41	50	26.0	79	23	ABP42461
42	50	26.0	496	23	AAU74443
43	50	26.0	613	21	AAU19339
44	50	26.0	674	23	AAU74445
45	50	26.0	714	23	AAU74441

## ALIGNMENTS

RESULT 1	AAU64657
ID	AAU64657 standard; Protein: 1152 AA.
XX	XX
AC	AAU64657;
XX	XX
DT	22-MAR-2001 (first entry)
XX	XX
DE	Human secreted protein BLAST search protein SEQ ID NO: 167.
XX	XX
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein.
XX	XX
OS	Homo sapiens.
XX	XX
FN	W0200077197-A1.
XX	XX
PD	21-DEC-2000.
XX	XX
PF	01-JUN-2000; 2000WO-US14934.
XX	XX
PR	11-JUN-1999; 99US-0138599.
XX	XX
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	XX
XX	(ROSE/) ROSEN C A.
PI	Rosen CA, Ruben SM, Komatsoulis GA;
XX	XX
DR	WPI; 2001-032312/04.
XX	XX

Amino acid sequenc  
ITG11 protein enc  
Novel human diagno  
12D3 antigen sequ  
Murine protein iso  
Rat protein isolat  
Rat protein isolat  
Murine A259. Mus  
Murine A259 polype  
Propionibacterium  
Novel human diagno  
Novel human diagno  
Novel human diagno  
12D3 antigen sequ  
Human polypeptide  
Human c-ski oncopr  
C-ski protein. Ga  
FB29 chicken c-ski  
FB29-encoded chick  
Chicken c-ski FB29  
Alternative versio  
Human polypeptide  
Human multi-protei  
Murine multi-copper  
Mouse protein sequ  
Potato sucrose pho  
NF-ATp. Mus sp.  
Human T lymphocyte  
Novel human secret  
Human ovarian anti  
Human protein sequ  
Amino acid sequenc  
Human protein sequ  
Human protein sequ

PT Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 XX  
 PS Disclosure; Page 543-546; 558pp; English.

CC The invention relates to the isolation of genes AAF32757-F32803 encoding  
 CC the human secreted proteins AAB64549-B64594. The sequence is a search  
 CC result from a BLASTX homology search. The genes and proteins are useful  
 CC for preventing, ameliorating or treating medical conditions, e.g. by  
 CC protein or gene therapy. The genes are isolated from a range of human  
 CC tissues disclosed in the specification. The nucleic acids, proteins,  
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment  
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer, and  
 CC other cancers of the adrenal gland, bone, bone marrow, breast,  
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound  
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
 CC and (f) infectious diseases such as viral, bacterial, fungal and  
 CC parasitic infections.

CC Sequence 1152 AA;  
 XX  
 XX

Query Match 100.0%; Score 192; DB 22; Length 1152;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-18;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEHPHYGTLPGPGPEFKTTLRVONLGCYVSGLI 35  
 DB 952 YEHPHYGTLPGPGPEFKTTLRVONLGCYVSGLI 986

RESULT 2  
 AAB64658  
 ID AAB64658 standard; Protein; 1152 AA.  
 AC AAB64658;  
 XX  
 XX 22-MAR-2001 (first entry)  
 DT  
 XX  
 XX Human secreted protein BLAST search protein SEQ ID NO: 168.  
 DE  
 XX Cytostatic; immunosuppressive; nocitropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnereary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 XX WO200077197-A1.  
 XX  
 XX 21-DEC-2000.  
 PD  
 XX 01-JUN-2000; 2000WO-US14934.  
 PF  
 XX 11-JUN-1999; 99US-0138599.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 PA  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 PI WPI; 2001-032312/04.  
 DR  
 XX WPI; 2001-032312/04.  
 XX  
 XX Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 XX  
 PS Disclosure; Page 547-551; 558pp; English.  
 XX  
 CC The invention relates to the isolation of genes AAF32757-F32803 encoding

CC the human secreted proteins AAB64549-B64594. The sequence is used as a  
 CC query sequence for doing BLASTX searches to identify homologous  
 CC sequences. The genes and proteins are useful for preventing,  
 CC ameliorating or treating medical conditions, e.g. by protein or gene  
 CC therapy. The genes are isolated from a range of human tissues disclosed  
 CC in the specification. The nucleic acids, proteins, antibodies and  
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the  
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal  
 CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's  
 CC disease, allergies, autoimmune haemolytic anaemia, autoimmune  
 CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,  
 CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular  
 CC disorders such as myocardial ischaemia; (d) wound healing; (e)  
 CC neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections.

CC Sequence 1152 AA;  
 XX  
 XX

Query Match 100.0%; Score 192; DB 22; Length 1152;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-18;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEHPHYGTLPGPGPEFKTTLRVONLGCYVSGLI 35  
 DB 952 YEHPHYGTLPGPGPEFKTTLRVONLGCYVSGLI 986

RESULT 3  
 AAY32242  
 ID AAY32242 standard; Protein; 1167 AA.  
 AC AAY32242;  
 XX  
 XX 15-FEB-2000 (first entry)  
 DT  
 XX  
 XX Human integrin subunit alpha-10.  
 DE  
 XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;  
 KW fibroblast; vaccine; marker.  
 XX  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 XX Key Location/Qualifiers  
 FH 1..22  
 FT /note= "signal peptide"  
 FT 23..1145  
 FT /note= "mature protein"  
 FT 23..1120  
 FT /note= "extracellular domain"  
 FT 1121..1145  
 FT /note= "transmembrane domain"  
 FT 1122..1167  
 FT /note= "cytoplasmic domain, specifically claimed  
 FT in Claim 21"  
 FT 162..359  
 FT /note= "I-domain"  
 FT 494..502  
 FT /note= "cation binding site motif"  
 FT 558..566  
 FT /note= "cation binding site motif"  
 FT 620..628  
 FT /note= "cation binding site motif"  
 FT 98  
 FT /note= "N-glycosylated"  
 FT 336  
 FT /note= "N-glycosylated"  
 FT 364  
 FT /note= "N-glycosylated"  
 FT 733  
 FT Modified-site

```

FT FT Modified-site /note= "N-glycosylated"
FT FT Modified-site 839
FT FT Modified-site /note= "N-glycosylated"
FT FT Modified-site 921
FT FT Modified-site /note= "N-glycosylated"
FT FT Modified-site 1018
FT FT Modified-site /note= "N-glycosylated"
FT FT Modified-site 1039
FT FT Modified-site /note= "N-glycosylated"
XX
XX MO9951639-A1.
XX
XX 14-OCT-1999.
XX
XX 31-MAR-1999; 99WO-SE00544.
XX
XX 02-APR-1998; 98SE-0001164.
XX 28-JAN-1999; 99SE-0000319.
XX
XX (ACT1-) ACTIVE BIOTECH AB.
XX
XX Lundgren-Akerlund E;
XX
XX WPI, 2000-052639/04.
XX N-PSDB; AAF34719.
XX
XX New isolated integrin subunit alpha-10, used as a marker or target
XX molecule for cells during development, regeneration and pathological
XX conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
XX inflammation.
XX
XX Claim 1; Fig 6; 90pp; English.
XX
XX This sequence represents novel human chondrocyte integrin subunit
XX alpha-10 (ISa10). A splice variant is given in AAF32243. The
XX invention relates to a recombinant or isolated integrin heterodimer
XX comprising the alpha10 subunit in association with subunit beta
XX (especially beta-1). The heterodimer and the subunit alpha-10 can
XX be used as markers or targets of all types of cells, e.g. of
XX chondrocytes, osteoblasts and fibroblasts. They can also be used:
XX for treating pathological conditions involving ISa10, such as
XX damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
XX for detecting the formation of cartilage during embryonal
XX development, physiological or therapeutic repair of cartilage,
XX or detecting regeneration of cartilage or chondrocytes during
XX transplantation of cartilage or chondrocytes; for selection and
XX analysis of for sorting, isolating or purification of chondrocytes
XX and for in vitro studies of differentiation of chondrocytes; and as
XX a target for anti-adhesive drugs or molecules in tendon, ligament,
XX skeletal muscle or other tissues where adhesion impairs the function
XX of the tissue (all claimed). ISa10 binding entities can be used to
XX determine the differentiation-state of cells during embryonic
XX development, angiogenesis or development of cancer, in pathological
XX conditions such as rheumatoid arthritis, osteoarthritis or cancer,
XX in tissue regeneration or in therapeutic and physiological repair of
XX of cartilage (claimed). A vaccine comprising the integrin
XX heterodimer or subunit alpha-10 is also claimed. ISa10
XX polynucleotides, vectors, host cells and methods of producing
XX recombinant ISa10 are also claimed.
XX
XX Sequence 1167 AA;
SQ
Query Match 100.0%; Score 192; DB 21; Length 1167;
Best Local Similarity 100.0%; Pred. No. 6.3e-18;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YEVAHYGTLPVGPGEFKTLRVQNLGCVVSGLI 35
DB 952 YEVAHYGTLPVGPGEFKTLRVQNLGCVVSGLI 986
RESULT 4
AAB64584

```

```

ID AAB64584 standard; Protein: 1167 AA.
XX
XX AAB64584;
AC
XX 22-MAR-2001 (first entry)
XX
XX Human secreted protein #37.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antifungal; hepatotropic; antidiabetic; antiinflammatory; anticancer;
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
XX
XX WO200077197-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US14934.
XX
XX 11-JUN-1999; 99US-0138599.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI, 2001-032312/04.
XX N-PSDB; AAF32793.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX Claim 11; Page 496-500; 558pp; English.
XX
XX Sequences AAB64549-B64594 represent the amino acid sequences of 47
XX human secreted proteins encoded by the genes AAF373757-F37803. The genes
XX and proteins are useful for preventing, ameliorating or treating medical
XX conditions, e.g. by protein or gene therapy. The genes are isolated from
XX a range of human tissues disclosed in the specification. The nucleic
XX acids, proteins, antibodies and (ant)agonists are useful in the
XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
XX Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Sequence 1167 AA;
SQ
Query Match 100.0%; Score 192; DB 22; Length 1167;
Best Local Similarity 100.0%; Pred. No. 6.3e-18;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YEVAHYGTLPVGPGEFKTLRVQNLGCVVSGLI 35
DB 952 YEVAHYGTLPVGPGEFKTLRVQNLGCVVSGLI 986
RESULT 5
AAY32243
ID AAY32243 standard; Protein: 1132 AA.
XX
XX AAY32243;
AC
XX 15-FEB-2000 (first entry)
XX
XX
XX
XX

```

DE Human integrin subunit alpha-10 splice variant.

KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;  
 KW fibroblast; vaccine; marker; splice variant.

OS Homo sapiens.

Key Location/Qualifiers

FT Peptide

1..22 /note="signal peptide"

FT Protein

23..1132 /note="mature protein"

MO9951639-A1.

14-OCT-1999.

31-MAR-1999; 99WO-SE00544.

02-APR-1998; 98SE-0001164.

28-JAN-1999; 99SE-0000319.

(ACT1-) ACTIVE BIOTECH AB.

Lundgren-Akerlund E;

WPI; 2000-052639/04.

N-PSDB; AAS24720.

PT New isolated integrin subunit alpha-10, used as a marker or target  
 molecule for cells during development, regeneration and pathological  
 conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 inflammation

Claim 1; Page 43-48; 90pp; English.

This sequence represents a splice variant of novel human  
 chondrocyte integrin subunit alpha-10 (ISa10). It is identical to  
 ISa10 (see AAI32242) except for deletion of amino acids 975-986. The  
 invention relates to a recombinant or isolated integrin heterodimer  
 comprising the alpha10 subunit in association with subunit beta  
 (especially beta-1). The heterodimer, subunit alpha-10 or splice  
 variant can be used as a marker or target of all types of cells, e.g.  
 of chondrocytes, osteoblasts and fibroblasts. They can also be used:  
 for treating pathological conditions involving ISa10, such as  
 damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
 for detecting the formation of cartilage during embryonal  
 development, physiological or therapeutic repairation of cartilage,  
 or detecting regeneration of cartilage or chondrocytes during  
 transplantation of cartilage or chondrocytes; for selection and  
 analysis or for sorting, isolating or purification of chondrocytes  
 and for in vitro studies of differentiation of chondrocytes; and as  
 a target for anti-adhesive drugs or molecules in tendon, ligament,  
 skeletal muscle or other tissues where adhesion impairs the function  
 of the tissue (all claimed). ISa10 binding entities can be used to  
 determine the differentiation-state of cells during embryonic  
 development, angiogenesis or development of cancer, in pathological  
 conditions such as rheumatoid arthritis, osteoarthritis or cancer,  
 in tissue regeneration or in therapeutic and physiological repairation  
 of cartilage (claimed). A vaccine comprising the integrin  
 heterodimer or subunit alpha-10 is also claimed. ISa10  
 polynucleotides, vectors, host cells and methods of producing  
 recombinant ISa10 are also claimed.

Sequence 1132 AA;

Query Match 74.0%; Score 142; DB 21; Length 1132;

Best Local Similarity 83.3%; Pred. No. 5 6e-11;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 YEVHPYGLPVGPGPEFKTTLRTNNASCIV 30

Db 952 YEVHPYGLPVGPGPEFKTTLRTNNASCIV 981

RESULT 6  
 AAU14231  
 AAU14231 standard; Protein; 1188 AA.

AAU14231;

24-OCT-2001 (first entry)

Human novel protein #102.

Human; novel protein; Antianaemic; osteopathic; antiinflammatory;  
 immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;  
 anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;  
 antibacterial; antiallergic; dermatological; haemostatic; antiaesthetic;  
 thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 tissue regeneration; immune disorder.

Homo sapiens.

WO200155437-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US02623.

25-JAN-2000; 2000US-0491404.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Dermanac RT;

WPI; 2001-451939/48.

N-PSDB; AAS22536.

PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
 PT nervous system disorders, and for regenerating bone and cartilage -  
 PS Example 4; Page 578-581; 894pp; English.

The invention relates to polynucleotides encoding novel human  
 proteins or their active domains. The polypeptides, polynucleotides and  
 antibodies raised against the polypeptides are used in a method of  
 treatment of a mammal and prevention of disorders caused by the aberrant  
 protein expression or activity. The polypeptides can be used as  
 molecular weight markers, food supplements, and in antibody production.  
 The polypeptides are used to identify compounds which bind to the  
 polypeptides. Polynucleotides of the invention are used as probes and  
 primers, for sequencing, for chromosome or gene mapping, in the  
 production of recombinant proteins, and in generating anti-sense DNA or  
 RNA and in gene therapy. Polypeptides of the invention can be used to  
 target drugs to a tumour, in assays to determine biological activity, to  
 raise antibodies/elicit an immune response, to determine quantitative  
 protein levels, as tissue markers, and to isolate receptors or ligands.  
 Polypeptides of the invention may also be useful in creating platelet  
 disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 ligament and/or nerve tissue, wound healing, treating burns, promoting  
 the proliferation, differentiation and survival of stem cells, as a  
 contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 fungal infection or from autoimmunity, cancer, allergy, asthma,  
 graft-versus-host disease, eczema, haemophilia, thrombosis,  
 anti-inflammatory diseases, nervous system disorders, and infection.  
 The present sequence represents a protein of the invention.

Sequence 1188 AA;

Query Match 35.9%; Score 69; DB 22; Length 1188;



PT diagnosis of fibrosis, e.g. of the liver  
 XX  
 PS Claim 8; Fig 1; 164pp; English.  
 XX  
 CC The present sequence is human integrin alpha subunit, A259. A259 is  
 CC homologous with the alpha1 and alpha10 integrin subunits and is  
 CC overexpressed in fibrosis. A259 is implicated in regulation of  
 CC proliferation, differentiation and/or function of many different cell  
 CC types. Inhibitors of A259 activity are useful for the treatment of liver  
 CC disease, particularly fibrosis, and also fibrosis in other organs  
 CC (specifically lung and kidney). In addition, A259 can be used for  
 CC treatment and prevention of cancer, osteoporosis, acute myeloid  
 CC leukaemia, HIV infection, and rheumatoid arthritis.  
 XX  
 SQ Sequence 1188 AA;  
 Query Match 35.9%; Score 69; DB 22; Length 1188;  
 Best Local Similarity 40.5%; Pred. No. 0.84;  
 Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;  
 QY 1 YEVPYGTLP--VGPGPEFKTTLRVONLGCYVSGLI 35  
 968 YEVPKNSLSERYDGIQPFSCIFRIQNLGLPFIHGM 1004  
 DB  
 RESULT 9  
 AAU10551 standard; Protein; 1188 AA.  
 XX  
 AC AAU10551;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Human A259 polypeptide.  
 XX  
 KW Human; A259; integrin alpha subunit; integrin alpha 10; secreted protein;  
 KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;  
 KW cartilage associated disorder; haematopoietic disorder; bone marrow;  
 KW immune related disease; apoptotic disorder; neuronal tissue disease;  
 KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;  
 KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;  
 KW antidiabetic; antianaemic; antiallergic; antiaesthetic; dermatological;  
 KW antidiabetic; anticonvulsant; antiparkinsonian.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..22  
 FT Peptide  
 FT /note= "Signal peptide"  
 FT 1..1141  
 FT Domain  
 FT /note= "Extracellular domain"  
 FT 23..1188  
 FT Protein  
 FT /note= "Mature human A259"  
 FT 37..90  
 FT Domain  
 FT /note= "Integrin alpha repeat domain. The specification  
 FT states that this domain exists in human A549"  
 FT 115..157  
 FT Domain  
 FT /note= "Integrin alpha repeat domain"  
 FT 164..345  
 FT Domain  
 FT /note= "I domain or Von Willebrand Factor type A domain"  
 FT 367..392  
 FT Domain  
 FT /note= "Integrin alpha repeat domain"  
 FT 421..472  
 FT Domain  
 FT /note= "Integrin alpha repeat domain. The specification  
 FT states that this domain exists in human A549"  
 FT 476..532  
 FT Domain  
 FT /note= "Integrin alpha repeat domain. The specification  
 FT states that this domain exists in human A549"  
 FT 538..593  
 FT Domain  
 FT /note= "Integrin alpha repeat domain. The specification  
 FT states that this domain exists in human A549"  
 FT 600..654  
 FT Domain  
 FT /note= "Integrin alpha repeat domain. The specification

FT states that this domain exists in human A549"  
 FT Domain  
 FT 1142..1164  
 FT /note= "Transmembrane domain"  
 FT Domain  
 FT 1165..1188  
 FT /note= "Cytoplasmic domain"  
 XX  
 PN WO200181414-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 27-APR-2001; 2001WO-US13516.  
 XX  
 PR 27-APR-2000; 2000US-0561263.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Pan Y, Lora J;  
 XX  
 DR WPI; 2002-041397/05.  
 DR N-PSDB; AAS16873.  
 XX  
 PT New A259 nucleic acids and polypeptides, which comprise integrin alpha  
 PT subunit, useful for diagnosing, preventing or treating e.g. liver  
 PT disease, kidney or lung fibrosis, cancers, blood disorders or immune  
 PT related diseases  
 XX  
 PS Claim 9; Fig 1; 168pp; English.  
 XX  
 CC The invention relates to human and murine A259 nucleic acid molecules  
 CC which encode secreted proteins with homology to integrin alpha subunits,  
 CC specifically to integrin alpha 10. The A259 polypeptide and nucleic acid  
 CC are useful for treating liver disease or fibrosis, particularly kidney  
 CC fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also  
 CC useful for diagnosing, preventing or treating cartilage and bone  
 CC associated disorders (such as bone cancer, achondroplasia, myeloma,  
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and  
 CC acute myeloid leukaemia, haemophilia, anaemia and thalassaemia), immune  
 CC related diseases (such as HIV, viral infections, cancers, T cell  
 CC autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.  
 CC asthma and psoriasis), apoptotic disorders (such as systemic lupus  
 CC erythematosus and insulin-dependent diabetes mellitus), diseases of the  
 CC neuronal tissues (such as epilepsy and muscular dystrophy) and  
 CC neurodegenerative diseases (such as Parkinson's disease and Huntington's  
 CC disease). This sequence represents the human A259 polypeptide.  
 XX  
 SQ Sequence 1188 AA;  
 Query Match 35.9%; Score 69; DB 23; Length 1188;  
 Best Local Similarity 40.5%; Pred. No. 0.84;  
 Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;  
 QY 1 YEVPYGTLP--VGPGPEFKTTLRVONLGCYVSGLI 35  
 968 YEVPKNSLSERYDGIQPFSCIFRIQNLGLPFIHGM 1004  
 DB  
 RESULT 10  
 AAB25590 standard; Protein; 1034 AA.  
 ID AAB25590  
 AC AAB25590;  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX  
 DE Protein encoded by human secreted protein gene #7 clone HOHBV69.  
 XX  
 KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;  
 KW antirheumatic; dermatological; antiproliferative; antiatherosclerotic;  
 KW anticancer; vulnary; antiviral; antibacterial; antifungal;  
 KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;  
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;  
 KW Crohn's disease; nephritis; hyperproliferative disorder;



KM cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;  
 KM melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.  
 OS Homo sapiens.  
 XX WO200029435-A1.  
 PN 25-MAY-2000.  
 PD 27-OCT-1999; 99WO-US25031.  
 PF 28-OCT-1998; 98US-0105971.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX  
 XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JU, Moore PA, Wei Y,  
 PI Greene JM;  
 DR WPI: 2000-387742/33.  
 XX  
 XX Isolated nucleic acid molecules encoding human secreted proteins are  
 PT for the prevention, amelioration and treatment of autoimmune,  
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,  
 PT wounds, and infectious diseases -  
 XX  
 XX Claim 1; Page 678-682; 803pp; English.  
 PS  
 XX The present invention relates to 12 secreted human proteins and the  
 CC nucleotide sequences encoding them. The polynucleotide sequences given  
 CC in AAB0606-A80623 encode the 12 secreted protein sequences given in  
 CC AAB5576-B5593. The human secreted proteins have various activities  
 CC dependant on the tissues in which they are expressed. Examples of the  
 CC activities of the proteins include: immunosuppressant;  
 CC anti-inflammatory; antiarthritic; antirheumatic; dermatological;  
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnary;  
 CC antiviral; antibacterial; and antifungal activity. The proteins,  
 CC polypeptides, agonists and antagonists may be used to treat prevent  
 CC and/or diagnose various disease, disorders and conditions examples of  
 CC which include: immune disorders e.g. Addison's disease, rheumatoid  
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders  
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;  
 CC hyperproliferative disorders such as paraproteinemia and purpura;  
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;  
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide  
 CC sequences may also be used in wound healing and the treatment of  
 CC infectious diseases. The human secreted protein gene #7 and protein  
 CC sequences are represented in sequences AAB0612 and AAB5582. Secreted  
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences  
 CC AAB0652-A80661 represent genes which are related to the secreted protein  
 CC gene#7.  
 XX  
 XX  
 SQ Sequence 1034 AA;  
 Query Match 30.7%; Score 59; DB 21; Length 1034;  
 Best Local Similarity 41.7%; Pred. No. 18;  
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
 Oy 12 GPGPEFKTLRVNIGCVVSGLI 35  
 Db 981 GIGPPFSCIFRIQNLGFPFHGM 1004  
 RESULT 11  
 AAB30929  
 ID AAB30929 standard; Protein; 1188 AA.  
 XX  
 XX AAB30929;  
 AC  
 XX  
 XX 02-APR-2001 (first entry)  
 DT  
 XX  
 XX Amino acid sequence of a human alpha1 integrin chain.  
 DE  
 XX  
 XX Human; integrin; alpha1 subunit; fibroblast; muscle cell; chondrocyte;  
 KM

KM osteoblast; stem cell; cell damage; muscle dystrophy; fibrosis;  
 KM wound healing; trauma; rheumatoid arthritis; osteoarthritis;  
 KM osteoporosis; cartilage damage; bone damage; cartilage.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..22  
 FT /note= "signal peptide"  
 FT 951..972  
 FT Region /note= "leucine zipper"  
 FT 1142..1164  
 FT Domain /note= "transmembrane domain"  
 XX  
 XX WO200075187-A1.  
 PN 14-DEC-2000.  
 PD 31-MAY-2000; 2000WO-SE01135.  
 PF 03-JUN-1999; 99SE-0002056.  
 PR (ACTI-) ACTIVE BIOTECH AB.  
 XX  
 XX Gullberg D;  
 PI  
 DR WPI: 2001-071061/08.  
 DR N-PsDB; AAC68671.  
 XX  
 XX Integrin subunit alpha 11 or integrin heterodimer comprising subunit  
 PT alpha 11 in association with subunit beta, useful for treating muscle  
 PT dystrophy, fibrosis, trauma, rheumatoid arthritis, and osteoarthritis -  
 PT  
 PS Disclosure: Fig 2a-c; 79pp; English.  
 XX  
 XX The present sequence represents a human integrin subunit, designated  
 CC alpha11. The alpha11 polynucleotide and polypeptide are useful as  
 CC markers of cell target molecules, such as fibroblasts, muscle cells,  
 CC chondrocytes, osteoblasts, mesenchymally derived cells or stem cells.  
 CC They are also used for determining the differential-stage of cells  
 CC during differentiation, development in pathological conditions, in  
 CC tissue regeneration, in transplantation or in therapeutic and  
 CC physiological repair of tissues. The pathological conditions involving  
 CC subunit alpha11 are selected from damage of cells, muscle dystrophy,  
 CC fibrosis, wound healing, trauma, rheumatoid arthritis, osteoarthritis  
 CC and osteoporosis, damage of cartilage and bone, and cartilage and bone  
 CC diseases. The polypeptide is useful for detecting the formation of  
 CC cartilage during embryonic development, for detecting physiological  
 CC therapeutic repair of cartilage and muscle, for selection and analysis,  
 CC or for sorting, isolating or purification of chondrocytes and muscle  
 CC cells, for detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes during transplantation of  
 CC cartilage or chondrocytes, respectively, or of muscle or muscle cells  
 CC during transplantation of muscle or muscle cells, respectively, and for  
 CC studies of differentiation or chondrocytes or muscle cells.  
 XX  
 XX  
 SQ Sequence 1188 AA;  
 Query Match 30.7%; Score 59; DB 22; Length 1188;  
 Best Local Similarity 41.7%; Pred. No. 21;  
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
 Oy 12 GPGPEFKTLRVNIGCVVSGLI 35  
 Db 981 GIGPPFSCIFRIQNLGFPFHGM 1004  
 RESULT 12  
 AAB25582  
 ID AAB25582 standard; Protein; 1189 AA.  
 XX  
 XX AAB25582;  
 AC  
 XX

21-NOV-2000 (first entry)  
ITGA11 protein encoded by human secreted protein gene #7

secreted protein; immunosuppressant; anti-inflammatory; antiatherosclerotic; antirheumatic; dermatological; antiproliferative; antiarteriosclerotic; anticancer; vulnery; antiviral; antibacterial; antifungal; autoimmune disorder; Addison's disease; rheumatoid arthritis; dermatitis; multiple sclerosis; inflammatory disorder; inflammatory bowel disease; Crohn's disease; nephritis; hyperproliferative disorder; cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer melanoma; lymphoma; wound healing; human chromosome 15 q23.3-23.

**Homo sapiens.**

W0200029435-A1.

25-MAY-2000.

27-OCT-1999; 99WO-US25031.

28-OCT-1998; 98US-0105971.

(HUMA-) HUMAN GENOME SCI INC.

Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y, Greene JM;

WPI; 2000-387742/33.

N-PSDB; AAA80612.

Isolated nucleic acid molecules encoding human secreted proteins are used for the prevention, amelioration and treatment of autoimmune, inflammatory, hyperproliferative and cardiovascular disorders, cancer wounds, and infectious diseases -

Claim 1; Figure 19A-F; 803pp; English.

The present invention relates to 12 secreted human proteins and the nucleotide sequences encoding them. The polynucleotide sequences given in AA80605-2-AB0623 encode the 12 secreted protein sequences given in AA825576-B25593. The human secreted proteins have various activities dependent on the tissues in which they are expressed. Examples of the activities of the proteins include: immunosuppressant; anti-inflammatory; anti-atheritis; anti-rheumatic; dermatological; antiproliferative; antiarteriosclerotic; anticancer; vulnery; antiviral; antibacterial; and antifungal activity. The proteins, polypeptides, agonists and antagonists may be used to treat prevent and/or diagnose various disease, disorders and conditions examples of which include: immune disorders e.g. Addison's disease, rheumatoid arthritis, dermatitis, and multiple sclerosis; inflammatory disorders e.g. inflammatory bowel disease, Crohn's disease and neuritis; hyperproliferative disorders such as parapsoriasis and purpura; cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis; cancer e.g. melanoma and lymphoma. The proteins and polynucleotide sequences may also be used in wound healing and the treatment of infectious diseases. The human secreted protein gene #7 and protein sequences are represented in sequences AA80612 and AA825582. Secreted protein gene #7 is located at position chromosome 15 q22.3-23. Sequences AA80652-2-AB0661 represent genes which are related to the secreted protein gene#7.

SQ Sequence 1189 AA.

Query Match	30.7%;	Score 59;	DB 21;	Length 1189;
Best Local Similarity	41.7%;	Pred. No. 21;		
Matches 10;	Conservative 5;	Mismatches 9;	Indels 0;	Gaps 0

```
Qy      12 GPGPEFKTLRVQNLGCYVWSGLI 35
          ||| | : ||| | : : :
Db      . 981 GIGPPFSCIFRIQNLGLFPPIHGIM 1004
```

RESULT 13  
ABG12949

AC ABG12949

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #12940

KM Human; chromosome mapping; gene mapping; gene therapy; forensic food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens

PN WO200175067-A2

PD 11-OCT-2001

PF 30-MAR-2001; 2001WO-US08631

PR 31-MAR-2000; 2000US-0540217

XX

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT

DR WPI; 2001-639362/73

DR N-PSDB; AAS77136

PT New isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity -

PS Claim 20; SEQ ID No 43308; 103pp; English

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://ftp.wipo.int/pub/published_pat_sequences).

**SQ Sequence 1189 AA.**

Query Match	30.7%;	Score 59;	DB 22;	Length 1189;
Best Local Similarity	41.7%;	Pred. NO. 21;		
Matches 10;	Conservative 5;	Mismatches 9;	Indels 0;	Gaps 0;

```

QY      12 GPPEPKTTLRVQNLGCYVWSGLI 35
      ||| | : ||| : : ||:
DB      981 GIGPPFSCIFRIQNLGLPFIHGMM 1004

```

RESULT 14  
AAR11069

```

ID AAR1069 standard; Protein; 347 AA.
XX
XX AAR1069;
XX
XX 23-MAY-1991 (first entry)
XX
XX 12D3 antigen sequence deduced from cDNA of pB1c1.
XX
XX Babesia; red water fever; cattle; antibodies.
XX
XX Babesia bovis.
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX Protein /label= signal sequence
XX 22..346
XX /label= mature protein
XX
XX EP417524-A.
XX
XX 20-MAR-1991.
XX
XX 23-AUG-1990; 90EP-0116165.
XX
XX 23-AUG-1989; 89AU-0005902.
XX
XX (CSIR ) COMMONWEALTH SCIENT ORG.
XX
XX Riddles PW, Aylward JH, Wright IG;
XX
XX WPI; 1991-081822/12.
XX
XX N-PSDB; AAQ10952.
XX
XX New antigen from Babesia for protective vaccine - used with
XX derived antibodies, DNA sequences and oligo:nucleotide probes,
XX for immuno-diagnosis.
XX
XX Claim 13; Fig 9; 40pp; English.
XX
XX The sequence was deduced from a clone isolated from a cDNA library
XX prep'd from bovine erythrocytes infected with Babesia bigemina.
XX The DNA can be used to construct a vector for expression of the
XX antigen which can be used in vaccines to protect cattle from red
XX water fever.
XX See also AAR1068.
XX
XX Sequence 347 AA;
XX
XX Query Match 29.2%; Score 56; DB 12; Length 347;
XX Best Local Similarity 44.7%; Pred. No. 15;
XX Matches 17; Conservative 2; Mismatches 13; Indels 6; Gaps 3;
OY 1 YEYHRY-GTLPVGP---GPEFKTLR--VQNLGCVWS 32
DB 67 YIVNRYIGRCPIGTTCGPEFDATLGTCA DSGCYTWS 104

```

## RESULT 15

```

ABR72288
ID ABR72288 standard; Protein; 545 AA.
XX
XX ABR72288;
XX
XX 04-APR-2002 (first entry)
XX
XX Murine protein isolated from skin cells SEQ ID NO: 500.
XX
XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
XX developmental defect; inflammatory disease; dermatological; vulnerary;
XX immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX
XX Mus sp.
XX

```

```

EN WO200190357-A1.
XX
XX 29-NOV-2001.
XX
XX 24-MAY-2001; 2001WO-NZ00099.
XX
XX 24-MAY-2000; 2000US-206650P.
XX
XX 25-JUL-2000; 2000US-221232P.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Watson JD, Strachan L, Sleeman M, Onrust R, Morrison JG, Kumble KD;
XX
XX WPI; 2002-122020/16.
XX
XX New polynucleotides and polypeptides encoded by the polynucleotides
XX isolated from skin cells, useful for treating skin wounds, cancers,
XX growth and developmental defects, inflammatory diseases, or for
XX modulating immune responses
XX
XX Claim 4; Page 305-306; 466pp; English.
XX
XX The present invention provides the protein and coding sequences of cDNAs
XX isolated from human, murine and rat skin cell libraries. The sequences
XX can be used in the development of therapeutic agents useful in the
XX treatment of skin diseases, including skin wounds, cancer, growth
XX defects, developmental defects and inflammatory diseases. The proteins
XX have important roles in the induction of hair growth, cell proliferation
XX and cell-cell interaction, in maintaining tissue integrity, in wound
XX healing and in modulating immune responses. The present sequence is a
XX polypeptide of the invention.
XX
XX Sequence 545 AA;
XX
XX Query Match 29.2%; Score 56; DB 23; Length 545;
XX Best Local Similarity 41.7%; Pred. No. 24;
XX Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
OY 12 GPGEFKTLRVQNLGCVWSGLI 35
DB 338 GIGPPFNCVFKNQNLGFPPIHGM 361

```

Search completed: July 16, 2003, 07:52:07  
Job time : 5.88889 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:55:38 ; Search time 2.5316 Seconds  
(without alignments)  
1639.577 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192  
Sequence: 1 YEHPHYGTLPVGPGEFRTLRVONLGCYVSGLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	30.7	1034	9	US-09-984-130-43
2	59	30.7	1034	9	US-09-836-353A-43
3	59	30.7	1189	9	US-09-984-130-35
4	59	30.7	1189	9	US-09-836-353A-35
5	57.5	28.9	3338	9	US-10-156-761-8464
6	56	28.2	548	9	US-09-866-050A-500
7	56	28.2	688	9	US-09-866-050A-524
8	56	28.2	696	9	US-09-866-050A-501
9	52	27.1	308	9	US-10-156-761-13191
10	52	27.1	728	10	US-09-753-831-2
11	51	26.6	846	10	US-09-376-045-2
12	51	26.6	908	10	US-09-376-045-6
13	51	26.6	1053	9	US-10-217-700-6
14	51	26.6	1054	10	US-09-376-045-4
15	50	26.0	737	9	US-10-106-698-4893
16	50	26.0	738	9	US-10-235-521-1
17	49	25.5	127	9	US-10-106-698-6472
18	49	25.5	127	9	US-09-912-697-35
19	48.5	25.3	1109	10	US-09-864-761-38086

20	48.5	25.3	255	9	US-10-125-540-427	Sequence 427, App
21	48.5	25.3	255	10	US-09-764-870-427	Sequence 427, App
22	48.5	25.3	470	9	US-10-156-761-14879	Sequence 14879, A
23	48	25.0	354	9	US-10-050-704-115	Sequence 115, App
24	48	25.0	383	9	US-10-050-704-239	Sequence 239, App
25	48	25.0	511	9	US-10-156-761-11125	Sequence 11125, A
26	48	25.0	668	10	US-09-918-909-2	Sequence 2, Appl1
27	48	25.0	1289	10	US-09-738-563-4	Sequence 4, Appl1
28	47.5	24.7	245	8	US-08-424-550B-4	Sequence 40, Appl1
29	47.5	24.7	270	9	US-09-738-636-5968	Sequence 5968, Ap
30	47.5	24.7	625	9	US-10-099-885-14	Sequence 34, Appl1
31	47.5	24.7	1065	9	US-09-771-161A-239	Sequence 239, App
32	47.5	24.7	1161	9	US-10-156-761-14082	Sequence 14082, A
33	47.5	24.7	2597	10	US-09-905-129-2	Sequence 2, Appl1
34	47.5	24.7	2597	10	US-09-905-129-10	Sequence 10, Appl1
35	47.5	24.7	2597	10	US-09-905-129-13	Sequence 13, Appl1
36	47.5	24.7	2597	10	US-09-991-630-2	Sequence 2, Appl1
37	47.5	24.7	2597	10	US-09-991-630-10	Sequence 10, Appl1
38	47.5	24.7	2597	10	US-09-991-630-13	Sequence 13, Appl1
39	47.5	24.7	2972	8	US-08-424-550B-387	Sequence 387, App
40	47	24.5	373	9	US-09-978-295A-59	Sequence 59, Appl
41	47	24.5	373	9	US-09-992-598-503	Sequence 503, App
42	47	24.5	373	9	US-09-978-687-59	Sequence 59, Appl
43	47	24.5	373	9	US-09-978-152A-59	Sequence 59, Appl
44	47	24.5	373	9	US-09-989-293A-503	Sequence 503, App
45	47	24.5	373	9	US-09-989-735-503	Sequence 503, App

#### ALIGNMENTS

RESULT 1  
US-09-984-130-43  
; Sequence 43, Application US/09984130  
; Publication No. US20030055231A1  
; GENERAL INFORMATION:  
; APPLICANT: NI et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: PFA4892  
; CURRENT APPLICATION NUMBER: US/09/984,130  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/243,792  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 09/836,353  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,407  
; PRIOR FILING DATE: 2000-04-19  
; PRIOR APPLICATION NUMBER: PCT/US99/25031  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR APPLICATION NUMBER: 60/105,971  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 149  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 43  
; LENGTH: 1034  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-130-43  
30.7% Score 59; DB 9; Length 1034;  
Best Local Similarity 41.7%; Pred. No. 11;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 12 GPGEFRTLRVONLGCYVSGLI 35  
Db 981 GIGPFSCIFRIOMLGLPIHGM 1004  
RESULT 2  
US-09-836-353A-43  
; Sequence 43, Application US/09836353A  
; Publication No. US20030129685A1  
; GENERAL INFORMATION:

APPLICANT: Ni et al.  
TITLE OF INVENTION: 12 Human Secreted Proteins  
FILE REFERENCE: PF489P1  
CURRENT APPLICATION NUMBER: US/09/836,353A  
CURRENT FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/198,407  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: PCT/US99/25031  
PRIOR FILING DATE: 1999-10-27  
PRIOR APPLICATION NUMBER: 60/105,971  
PRIOR FILING DATE: 1998-10-28  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 43  
LENGTH: 1034  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-836-353A-43

Query Match 30.7%; Score 59; DB 9; Length 1034;  
Best Local Similarity 41.7%; Pred. No. 11;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGPEFKTTLRVQNLGCVVSGLI 35  
Db 981 GIGPFSCTFRIONLGLFPIHGIM 1004

RESULT 3  
US-09-984-130-35  
Sequence 35, Application US/09984130  
Publication No. US20030055231A1  
GENERAL INFORMATION:  
APPLICANT: Ni et al.  
TITLE OF INVENTION: 12 Human Secreted Proteins  
FILE REFERENCE: PF489P2  
CURRENT APPLICATION NUMBER: US/09/984,130  
CURRENT FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: 60/243,792  
PRIOR FILING DATE: 2000-10-30  
PRIOR APPLICATION NUMBER: 09/836,353  
PRIOR FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/198,407  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: PCT/US99/25031  
PRIOR FILING DATE: 1999-10-27  
PRIOR APPLICATION NUMBER: 60/105,971  
PRIOR FILING DATE: 1998-10-28  
NUMBER OF SEQ ID NOS: 149  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 35  
LENGTH: 1189  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-984-130-35

Query Match 30.7%; Score 59; DB 9; Length 1189;  
Best Local Similarity 41.7%; Pred. No. 13;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGPEFKTTLRVQNLGCVVSGLI 35  
Db 981 GIGPFSCTFRIONLGLFPIHGIM 1004

RESULT 4  
US-09-836-353A-35  
Sequence 35, Application US/09836353A  
Publication No. US20030129685A1  
GENERAL INFORMATION:  
APPLICANT: Ni et al.  
TITLE OF INVENTION: 12 Human Secreted Proteins  
FILE REFERENCE: PF489P1

CURRENT APPLICATION NUMBER: US/09/836,353A  
CURRENT FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/198,407  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: PCT/US99/25031  
PRIOR FILING DATE: 1999-10-27  
PRIOR APPLICATION NUMBER: 60/105,971  
PRIOR FILING DATE: 1998-10-28  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 35  
LENGTH: 1189  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-836-353A-35

Query Match 30.7%; Score 59; DB 9; Length 1189;  
Best Local Similarity 41.7%; Pred. No. 13;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGPEFKTTLRVQNLGCVVSGLI 35  
Db 981 GIGPFSCTFRIONLGLFPIHGIM 1004

RESULT 5  
US-10-156-761-8464  
Sequence 8464, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 8464  
LENGTH: 3338  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-8464

Query Match 29.9%; Score 57.5; DB 9; Length 3338;  
Best Local Similarity 56.5%; Pred. No. 67;  
Matches 13; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

Qy 8 TLVPQGPPEFKTTLRVQNLGCVV 30  
Db 2091 TLP-GPDPFSPARLRKGLGLOV 2112

RESULT 6  
US-09-866-050A-500  
Sequence 500, Application US/09866050A  
Publication No. US20030040477A1  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew  
APPLICANT: Ornst, Rene  
APPLICANT: Murison, James G.  
APPLICANT: Kumble, Krishanand D.  
TITLE OF INVENTION: Compositions Isolated From Skin Cells

```

; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 500
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-500

Query Match          29.2%; Score 56; DB 9; Length 545;
Best Local Similarity 41.7%; Pred. No. 14;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GGPPEFKTLRVONLGCYVSGLI 35
Db 338 GIGPPENCVFVKVQNLGFFPIHGV 361

RESULT 7
US-09-866-050A-624
; Sequence 624, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 624
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-624

Query Match          29.2%; Score 56; DB 9; Length 688;
Best Local Similarity 41.7%; Pred. No. 18;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GGPPEFKTLRVONLGCYVSGLI 35
Db 481 GIGPPENCVFVKVQNLGFFPIHGV 504

RESULT 8
US-09-866-050A-501
; Sequence 501, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
```

```

; LENGTH: 696
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-501

Query Match          29.2%; Score 56; DB 9; Length 696;
Best Local Similarity 41.7%; Pred. No. 18;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GGPPEFKTLRVONLGCYVSGLI 35
Db 481 GIGPPENCVFVKVQNLGFFPIHGV 504

RESULT 9
US-10-156-761-13191
; Sequence 13191, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13191
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13191

Query Match          27.1%; Score 52; DB 9; Length 308;
Best Local Similarity 38.7%; Pred. No. 26;
Matches 12; Conservative 6; Mismatches 9; Indels 4; Gaps 1;

Oy 5 PYGTLPVGPPEFKTLRVONLGCYVSGLI 35
Db 105 PYGTADG---FTQFGVNHIGHFALTGL 131

RESULT 10
US-09-753-831-2
; Sequence 2, Application US/09753831
; Patent No. US20020137683A1
; GENERAL INFORMATION:
; APPLICANT: Hogan, Kevin T.
; APPLICANT: Ross, Mark W.
; TITLE OF INVENTION: C-SKI Oncogene-Derived Peptides for Prevention,
; FILE REFERENCE: 26747-27
; CURRENT APPLICATION NUMBER: US/09/753,831
; CURRENT FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: U.S. 60/174296
; PRIOR FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 728
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-753-831-2

Query Match          27.1%; Score 52; DB 10; Length 728;
```

Best Local Similarity 35.3%; Pred. No. 69;  
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGPEF-----KTLRVQNLGCVVSG 33  
Db 84 PVLPGPFMPSPDRSTERCETVEGETISCFVVG 117

RESULT 11

US-09-376-045-2  
; Sequence 2, Application US/09376045  
; Patent No. US20020019998A1  
; GENERAL INFORMATION:  
; APPLICANT: Aventis CropScience GmbH  
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF  
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION  
; FILE REFERENCE: 514413-3772  
; CURRENT APPLICATION NUMBER: US/09/376,045  
; CURRENT FILING DATE: 1999-08-17  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 846  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-09-376-045-2

Query Match 26.6%; Score 51; DB 10; Length 846;  
Best Local Similarity 35.5%; Pred. No. 1.1e+02;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGLPVGPGEFFKTLRVQNLGCVV 31  
Db 704 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 732

RESULT 12

US-09-376-045-6  
; Sequence 6, Application US/09376045  
; Patent No. US20020019998A1  
; GENERAL INFORMATION:  
; APPLICANT: Aventis CropScience GmbH  
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF  
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION  
; FILE REFERENCE: 514413-3772  
; CURRENT APPLICATION NUMBER: US/09/376,045  
; CURRENT FILING DATE: 1999-08-17  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-09-376-045-6

Query Match 26.6%; Score 51; DB 10; Length 908;  
Best Local Similarity 35.5%; Pred. No. 1.2e+02;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGLPVGPGEFFKTLRVQNLGCVV 31  
Db 726 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 13

US-10-217-700-6  
; Sequence 6, Application US/10217700  
; Publication No. US20030070191A1  
; GENERAL INFORMATION:  
; APPLICANT: Haigler, Candace H.  
; APPLICANT: Holdaday, A. Scott  
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED  
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE

FILE REFERENCE: 201304/1000  
; CURRENT APPLICATION NUMBER: US/10/217,700  
; CURRENT FILING DATE: 2002-08-12  
; EARLIER APPLICATION NUMBER: 09/394,272  
; EARLIER FILING DATE: 1999-09-10  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 1053  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-10-217-700-6

Query Match 26.6%; Score 51; DB 9; Length 1053;  
Best Local Similarity 35.5%; Pred. No. 1.5e+02;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGLPVGPGEFFKTLRVQNLGCVV 31  
Db 911 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 939

RESULT 14

US-09-376-045-4  
; Sequence 4, Application US/09376045  
; Patent No. US20020019998A1  
; GENERAL INFORMATION:  
; APPLICANT: Aventis CropScience GmbH  
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF  
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION  
; FILE REFERENCE: 514413-3772  
; CURRENT APPLICATION NUMBER: US/09/376,045  
; CURRENT FILING DATE: 1999-08-17  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 1054  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-09-376-045-4

Query Match 26.6%; Score 51; DB 10; Length 1054;  
Best Local Similarity 35.5%; Pred. No. 1.5e+02;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGLPVGPGEFFKTLRVQNLGCVV 31  
Db 912 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 940

RESULT 15

US-10-106-698-4893  
; Sequence 4893, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: P000521  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 4893  
; LENGTH: 737  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:49:17 ; Search time 1.37834 Seconds  
(without alignments)  
747.132 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192  
Sequence: 1 YEHPYGTLPVGPDPFRTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pap:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pap:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pap:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pap:\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pap:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	27.1	750	4 US-08-202-841A-2	Sequence 2, Appli
2	51	26.6	846	1 US-08-356-354-2	Sequence 2, Appli
3	51	26.6	846	1 US-08-778-656-2	Sequence 2, Appli
4	51	26.6	890	1 US-08-145-006C-5	Sequence 5, Appli
5	51	26.6	890	5 PCT-US94-00545-5	Sequence 5, Appli
6	51	26.6	908	1 US-08-356-354-6	Sequence 6, Appli
7	51	26.6	908	2 US-08-778-656-6	Sequence 6, Appli
8	51	26.6	1054	1 US-08-356-354-4	Sequence 6, Appli
9	51	26.6	1054	2 US-08-778-656-4	Sequence 6, Appli
10	50	26.0	738	4 US-08-989-385-1	Sequence 4, Appli
11	48	25.0	668	4 US-09-697-367-2	Sequence 2, Appli
12	48	25.0	1289	1 US-07-876-280-4	Sequence 4, Appli
13	48	25.0	1289	1 US-07-673-772-4	Sequence 4, Appli
14	48	25.0	1289	1 US-08-063-170-4	Sequence 4, Appli
15	48	25.0	1289	1 US-08-158-232-4	Sequence 4, Appli
16	48	25.0	1289	1 US-08-304-626-4	Sequence 4, Appli
17	48	25.0	1289	1 US-08-316-301A-4	Sequence 4, Appli
18	48	25.0	1289	2 US-08-611-928-4	Sequence 4, Appli
19	48	25.0	1289	3 US-09-173-891-4	Sequence 4, Appli
20	48	25.0	1289	4 US-09-076-137-4	Sequence 4, Appli
21	48	25.0	1289	5 PCT-US92-03624-4	Sequence 4, Appli
22	48	25.0	1289	6 5281530-3	Sequence 4, Appli
23	48	25.0	1289	6 5426049-4	Sequence 4, Appli
24	47.5	24.7	245	4 US-08-469-260A-40	Sequence 40, Appli
25	47.5	24.7	1065	4 US-09-412-545-2	Sequence 2, Appli
26	47.5	24.7	1065	4 US-08-469-260A-387	Sequence 387, Appli
27	47	24.5	1057	4 US-08-853-948B-2	Sequence 2, Appli

28	47	24.5	1057	4 US-09-697-367-23	Sequence 23, Appli
29	46	24.0	33	1 US-08-257-528B-84	Sequence 84, Appli
30	46	24.0	33	1 US-08-460-602A-84	Sequence 84, Appli
31	46	24.0	33	1 US-08-463-966A-84	Sequence 84, Appli
32	46	24.0	33	1 US-08-465-217A-84	Sequence 84, Appli
33	46	24.0	33	2 US-08-464-329A-84	Sequence 84, Appli
34	46	24.0	33	2 US-08-462-507A-84	Sequence 84, Appli
35	46	24.0	33	2 US-08-467-881A-84	Sequence 84, Appli
36	46	24.0	34	4 US-09-141-833-11	Sequence 11, Appli
37	46	24.0	1045	4 US-08-553-436A-6	Sequence 6, Appli
38	46	24.0	1193	4 US-09-227-725A-4	Sequence 4, Appli
39	45.5	23.7	33	1 US-08-257-528B-83	Sequence 83, Appli
40	45.5	23.7	33	1 US-08-460-602A-83	Sequence 83, Appli
41	45.5	23.7	33	1 US-08-463-966A-83	Sequence 83, Appli
42	45.5	23.7	33	1 US-08-465-217A-83	Sequence 83, Appli
43	45.5	23.7	33	2 US-08-464-329A-83	Sequence 83, Appli
44	45.5	23.7	33	2 US-08-462-507A-83	Sequence 83, Appli
45	45.5	23.7	33	2 US-08-467-881A-83	Sequence 83, Appli

## ALIGNMENTS

RESULT 1  
US-08-202-841A-2  
Sequence 2, Application US/08202841A  
Patent No. 6218596  
GENERAL INFORMATION:  
APPLICANT: Hughes, Stephen H.  
APPLICANT: Sutlive, Pirmod  
APPLICANT: Pursel, Vernon  
TITLE OF INVENTION: Enhancement of Musculature in Animals  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,841A  
FILING DATE: 23-FEB-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/620,415  
FILING DATE: 03-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,449  
FILING DATE: 02-JUL-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/373,864  
FILING DATE: 30-JUN-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Alicea, Hector A.  
REGISTRATION NUMBER: 40,891  
REFERENCE/DOCKET NUMBER: 015280-170300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 750 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Modified-site

LOCATION: 373  
OTHER INFORMATION: /product= "OTHER"  
OTHER INFORMATION: /note= "Xaa = Trp in c-ski";  
OTHER INFORMATION: Xaa = Arg in v-ski"  
US-08-202-841A-2

Query Match 27.1%; Score 52; DB 4; Length 750;  
Best Local Similarity 35.3%; Pred. No. 21;  
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVONLGCYVVG 33  
DB 67 PVMPGFMPSPDRSTERCETITGETISCYVVG 100

RESULT 2  
US-08-356-354-2  
Sequence 2, Application US/08356354

Patent No. 5767365

GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe

TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE

NUMBER OF SEQUENCES: 6 PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen

STREET: 1180 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10036-8403

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/356,354

FILING DATE: 20-DEC-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/EP93/01605

FILING DATE: 22-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P42 20 758.4

FILING DATE: 24-JUN-1992

ATTORNEY/AGENT INFORMATION:

NAME: Meilman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-105

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 846 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-356-354-2

Query Match 26.6%; Score 51; DB 1; Length 846;

Best Local Similarity 35.5%; Pred. No. 33;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTLRVONLGCYV 31  
DB 704 FKVKCKPGTVP--PSKELRKVMRIQALRCHAV 732

RESULT 3  
US-08-778-656-2

Sequence 2, Application US/08778656

Patent No. 5976869

GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe

TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE

NUMBER OF SEQUENCES: 6 PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen

STREET: 1180 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10036-8403

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/778,656

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/356,354

FILING DATE: 20-DEC-1994

APPLICATION NUMBER: US PCT/EP93/01605

FILING DATE: 22-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P42 20 758.4

FILING DATE: 24-JUN-1992

ATTORNEY/AGENT INFORMATION:

NAME: Meilman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-105

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 846 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-778-656-2

Query Match 26.6%; Score 51; DB 2; Length 846;

Best Local Similarity 35.5%; Pred. No. 33;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTLRVONLGCYV 31  
DB 704 FKVKCKPGTVP--PSKELRKVMRIQALRCHAV 732

RESULT 4  
US-08-145-006C-5

Sequence 5, Application US/08145006C

Patent No. 5656452

GENERAL INFORMATION:

APPLICANT: Rao, Anjana

APPLICANT: Hogan, Patrick Gerald

APPLICANT: McCaffrey, Patricia

APPLICANT: Jain, Jyoti

TITLE OF INVENTION: NF-ATP, A T LYMPHOCYTE

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESSES:

ADDRESS: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/145,006C  
FILING DATE: October 29, 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/017,052  
FILING DATE: February 11, 1993  
APPLICATION NUMBER: 08/006,067  
FILING DATE: January 15, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Frazer, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 04590/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 890  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-145-006C-5

Query Match 26.6%; Score 51; DB 1; Length 890;  
Best Local Similarity 75.0%; Pred. No. 35;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 HPYGLPYGPGP 15  
|||  
Db 805 HPLGTCVPLPGP 816

RESULT 5  
PCT-US94-00545-5  
Sequence 5, Application PC/TUS9400545  
GENERAL INFORMATION:  
APPLICANT: Rao, Anjana  
APPLICANT: Hogan, Patrick Gerald  
APPLICANT: McCaffrey, Patricia  
APPLICANT: Jain, Jyugu  
TITLE OF INVENTION: NF-ATP, A T LYMPHOCYTE  
TITLE OF INVENTION: DNA-BINDING PROTEIN  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/00545  
FILING DATE: 18-JAN-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/145,006  
FILING DATE: October 29, 1993  
APPLICATION NUMBER: 08/017,052  
FILING DATE: February 11, 1993

APPLICATION NUMBER: 08/006,067  
FILING DATE: January 15, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Frazer, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 04590/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 890  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US94-00545-5

Query Match 26.6%; Score 51; DB 5; Length 890;  
Best Local Similarity 75.0%; Pred. No. 35;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 HPYGLPYGPGP 15  
|||  
Db 805 HPLGTCVPLPGP 816

RESULT 6  
US-08-356-354-6  
Sequence 6, Application US/08356354  
Patent No. 5767365  
GENERAL INFORMATION:  
APPLICANT: SONNEMAU, Uwe  
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE  
TITLE OF INVENTION: PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Oestlenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10036-8403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,354  
FILING DATE: 20-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/EP93/01605  
FILING DATE: 22-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P42 20 758.4  
FILING DATE: 24-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weilman, Edward A.  
REGISTRATION NUMBER: 24,735  
REFERENCE/DOCKET NUMBER: P/951-105  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
TELEX: 236925  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 908 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-356-354-6

Query Match 26.6%; Score 51; DB 1; Length 908;  
Best Local Similarity 35.5%; Pred. No. 36;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTTLRVONIGCYV 31  
DB 726 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 7  
US-08-778-656-6  
Sequence 6, Application US/08778656  
Patent No. 5976869  
GENERAL INFORMATION:  
APPLICANT: SONNEWALD, Uwe  
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE  
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10036-8403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/778,656  
FILING DATE:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/356,354  
FILING DATE: 20-DEC-1994  
APPLICATION NUMBER: US PCT/EP93/01605  
FILING DATE: 22-JUN-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: DE P42 20 758.4  
FILING DATE: 24-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weilman, Edward A.  
REGISTRATION NUMBER: 24,735  
REFERENCE/DOCKET NUMBER: P/951-105  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
TELEX: 236925  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 908 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-778-656-6

Query Match 26.6%; Score 51; DB 2; Length 908;  
Best Local Similarity 35.5%; Pred. No. 36;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTTLRVONIGCYV 31  
DB 726 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 8  
US-08-356-354-4  
Sequence 4, Application US/08356354  
Patent No. 5767365  
GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe  
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE  
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10036-8403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,354  
FILING DATE: 20-DEC-1994  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US PCT/EP93/01605  
FILING DATE: 22-JUN-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: DE P42 20 758.4  
FILING DATE: 24-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weilman, Edward A.  
REGISTRATION NUMBER: 24,735  
REFERENCE/DOCKET NUMBER: P/951-105  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
TELEX: 236925  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1054 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-356-354-4

Query Match 26.6%; Score 51; DB 1; Length 1054;  
Best Local Similarity 35.5%; Pred. No. 43;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTTLRVONIGCYV 31  
DB 912 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 940

RESULT 9  
US-08-778-656-4  
Sequence 4, Application US/08778656  
Patent No. 5976869  
GENERAL INFORMATION:  
APPLICANT: SONNEWALD, Uwe  
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE  
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10036-8403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

```

; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1:

```

```

;      COUNTRY:  USA
;      ZIP:      32606
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk

```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/876,280  
FILING DATE: 19920430  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/S 104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1289 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
INDIVIDUAL ISOLATE: PS17  
IMMEDIATE SOURCE:  
CLONE: E. coli NMS22 (pmc 1628) NRRL B-18652  
US-07-876-280-4

Query Match 25.0%; Score 48; DB 1; Length 1289;  
Best Local Similarity 38.7%; Pred. No. 1.5e+02;  
Matches 12; Conservative 7; Mismatches 8; Indels 4; Gaps 2;

QY 8 TLPEV-GPGPERKTLRVQ--NLGCVVSGL 34  
DB 743 TIPISGKDFNTLEIODIVSIDIFVGSGL 773

RESULT 13  
US-07-675-772-4  
Sequence 4, Application US/07675772  
Patent No. 5262399  
GENERAL INFORMATION:  
APPLICANT: Hickie, Leslie A.  
APPLICANT: Sick, August J.  
APPLICANT: Schwab, George E.  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Payne, Jewel M.  
TITLE OF INVENTION: No. 5262399el Compositions and Methods for the Control of  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ROMAN SALIWANCHIK  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/675,772  
FILING DATE: 19910327  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SALIWANCHIK, ROMAN  
REGISTRATION NUMBER: 21,023  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1289 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
INDIVIDUAL ISOLATE: PS17  
IMMEDIATE SOURCE:  
LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF KENNETH NARVA  
CLONE: 17B  
US-07-675-772-4

Query Match 25.0%; Score 48; DB 1; Length 1289;  
Best Local Similarity 38.7%; Pred. No. 1.5e+02;  
Matches 12; Conservative 7; Mismatches 8; Indels 4; Gaps 2;

QY 8 TLPEV-GPGPERKTLRVQ--NLGCVVSGL 34  
DB 743 TIPISGKDFNTLEIODIVSIDIFVGSGL 773

RESULT 14  
US-08-063-170-4  
Sequence 4, Application US/08063170  
Patent No. 5350576  
GENERAL INFORMATION:  
APPLICANT: Kim, Leo  
APPLICANT: Schwab, George E.  
TITLE OF INVENTION: Compositions and Methods for Inducing an Immune  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID R. SALIWANCHIK  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/063,170  
FILING DATE: 19930517  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,141  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/759,248  
FILING DATE: 13-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: SALIWANCHIK, DAVID R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/S 103.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1289 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein



HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: PS17  
INDIVIDUAL ISOLATE: PS17b  
US-08-063-170-4

Query Match 25.0%; Score 48; DB 1; Length 1289;  
Best Local Similarity 38.7%; Pred. No. 1.5e+02;  
Matches 12; Conservative 7; Mismatches 8; Indels 4; Gaps 2;

Qy 8 TLVP-GPGPERKTLRVQ--NLGCVVSGL 34  
Db 743 TIPPGSGKDFNTLEIIDIIVSIDIIVSGSL 773

RESULT 15  
US-08-158-232-4  
Sequence 4, Application US/08158232  
Patent No. 5596071  
GENERAL INFORMATION:

APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Vick, Heidi Jane  
APPLICANT: Foncerada, Luis  
APPLICANT: Schepf, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active  
NUMBER OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51

## CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/158,232  
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/887,980

FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/797,645

FILING DATE: 25-NOV-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/703,977

FILING DATE: 22-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: M/SCJ104.C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1289 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
STRAIN: PS17  
INDIVIDUAL ISOLATE: PS17b  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC1628) NRRL B-18652  
US-08-158-232-4

Query Match 25.0%; Score 48; DB 1; Length 1289;  
Best Local Similarity 38.7%; Pred. No. 1.5e+02;  
Matches 12; Conservative 7; Mismatches 8; Indels 4; Gaps 2;

Qy 8 TLVP-GPGPERKTLRVQ--NLGCVVSGL 34  
Db 743 TIPPGSGKDFNTLEIIDIIVSIDIIVSGSL 773

Search completed: July 16, 2003, 07:57:48  
Job time : 2.37834 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:21:09 ; Search time 39 Seconds  
(without alignments)  
86,274 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192  
Sequence: 1 YEVHPYGTLPVGPPEFKTLRVONLGCYVWSGLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 7751

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	19.3	35	2	PC2297
2	36	18.8	2	2	S05414
3	35.5	18.5	33	2	S13863
4	35	18.2	33	2	D61563
5	35	18.2	33	2	A54257
6	34.5	18.0	35	2	B36912
7	34	17.7	35	2	PC2295
8	33	17.2	10	2	H60588
9	33	17.2	30	2	T69492
10	32	16.7	22	2	P00070
11	32	16.7	30	2	S74112
12	32	16.7	35	2	PC2296
13	32	16.7	35	2	B41161
14	31.5	16.4	35	2	AD2525
15	31	16.1	19	2	S68394
16	31	16.1	20	2	A36016
17	31	16.1	24	2	T42441
18	31	16.1	24	2	B30609
19	31	16.1	28	2	A31859
20	31	16.1	28	2	JN0366
21	31	16.1	34	2	C31514
22	30.5	15.9	35	2	PC2294
23	30.5	15.9	35	2	S18264
24	30	15.6	18	2	C32537
25	30	15.6	20	2	PL0161
26	30	15.6	20	2	D32537
27	30	15.6	27	2	C54257
28	30	15.6	28	2	S41774
29	30	15.6	28	2	I48349

30	30	15.6	33	2	A95047	hypothetical prote
31	30	15.6	34	2	H95019	hypothetical prote
32	30	15.6	35	2	PC2293	V3 domain peptide
33	29.5	15.4	27	2	A24487	aldose 1-epimerase
34	29	15.1	15	2	S08282	cytochrome P450K-2
35	29	15.1	22	2	A39269	LX-1 tumor antigen
36	29	15.1	27	2	S64717	formin binding pro
37	29	15.1	30	2	A34874	transforming prote
38	29	15.1	35	2	PC2298	V3 domain peptide
39	28.5	14.8	33	2	A60507	pepsin I (EC 3.4.2
40	28	14.6	19	2	S69153	Neb-collagenase
41	28	14.6	20	2	S65399	immunodeficiency v
42	28	14.6	20	2	P00033	aldose 1-epimerase
43	28	14.6	22	2	H30608	ig kappa chain V-I
44	28	14.6	25	2	S36378	ig heavy chain V r
45	28	14.6	26	2	G30608	ig kappa chain V-I

## ALIGNMENTS

RESULT 1  
PC2297  
V3 domain peptide P4611 - human immunodeficiency virus type 1 (fragment)  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 21-Feb-1995 #sequence\_revision 05-Apr-1995 #text\_change 07-May-1999  
C:Accession: PC2297  
R:Sherefa, K.; Soenherborg, A.; Steinberg, J.; Saelberg, M.  
Biochem. Biophys. Res. Commun. 205:1658-1664, 1994  
A>Title: Rapid grouping of HIV-1 infection in subtypes A to E by V3 peptide serotyping a  
A:Reference number: PC2291, MUID:95110306; PMID:7811250  
A:Accession: PC2297  
A:Molecule type: protein  
A:Residues: 1-35 <SHE>  
C:Superfamily: type E retrovirus env polyprotein

Query Match 19.3% Score 37; DB 2; Length 35;  
Best Local Similarity 46.2% Pred. No. 1.7e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 TLVPGPPEFKTT 20  
DB 11 SIPIGGRALYTT 23

RESULT 2  
S05414  
steryl-sulfatase (EC 3.1.6.2), microsomal - rat (fragment)  
N:Alternate names: arylsulfatase C  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 28-Apr-1993  
C:Accession: S05414  
R:Kawano, U.I.; Kotani, T.; Ohteki, S.; Minamino, N.; Matsuo, H.; Ohnuma, T.; Aikawa, E.  
Biochem. Biophys. Acta 997, 199-205, 1989  
A>Title: Characterization of rat and human steroid sulfatases.  
A:Reference number: S05414; MUID:89352671; PMID:2765556  
A:Accession: S05414  
A:Molecule type: protein  
A:Residues: 1-26 <KAW>  
C:Keywords: sulfuric ester hydrolase

Query Match 18.8% Score 36; DB 2; Length 26;  
Best Local Similarity 40.9% Pred. No. 1.7e+02;  
Matches 9; Conservative 2; Mismatches 7; Indels 4; Gaps 1;

QY 12 GPGPEF---KTLRVONLGCY 29  
DB 4 GPGPFILIMADLIGIGLXGY 25

RESULT 3  
S13863  
dihydrolipoamide dehydrogenase (EC 1.8.1.4) - Trypanosoma cruzi (fragment)



OY 7 GTLPVPGP 14  
 |||||  
 Db 1 GTLPRTSG 8

## RESULT 9

169492  
 gene aeg-46.5 protein - Escherichia coli  
 C:Species: Escherichia coli  
 C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 08-Oct-1999  
 C:Accession: 169492  
 R:Choe, M.; Reznikoff, W.S.  
 J. Bacteriol. 175, 1165-1172, 1993  
 A:Title: Identification of the regulatory sequence of an anaerobically expressed locus aeg-46.5  
 A:Reference number: 154984; MUID:93163046; PMID:8432709  
 A:Accession: 169492  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-30 <RES>  
 A:Cross-references: GB:S54754; NID:9265388; PID:AA825330.1; PID:9265389

Query Match 17.7%; Score 32; DB 2; Length 30;  
 Best Local Similarity 31.4%; Pred. No. 5.4e+02;  
 Matches 11; Conservative 2; Mismatches 4; Indels 18; Gaps 2;

OY 3 VHPYG-----TLVPGPPEPKTTLRYQNL 26  
 :|||:|||||  
 Db 1 MHPVAGSLVAGAKPVTVSVRGP-----VMNL 28

## RESULT 10

P00070  
 T-cell receptor beta chain (BTB15) - bovine (fragment)  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Feb-1995  
 C:Accession: P00070  
 R:Tanaka, A.; Ishiguro, N.; Shinagawa, M.  
 submitted to JIPID, May 1990  
 A:Description: Sequence analysis of bovine T-cell receptor beta chain genes.  
 A:Reference number: J00472  
 A:Accession: P00070  
 A:Molecule type: mRNA  
 A:Residues: 1-22 <TRAN>  
 A:Experimental source: T cell  
 C:Genetics:  
 A:Gene: BTB15  
 C:Keywords: receptor

Query Match 16.7%; Score 32; DB 2; Length 22;  
 Best Local Similarity 66.7%; Pred. No. 5.3e+02;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 6 YGTLPVPGP 14  
 |||||  
 Db 8 YGELHFGPG 16

## RESULT 11

S74112  
 proline-rich antibacterial protein - green crab (fragment)  
 C:Species: Carcinus maenas (green crab, common shore crab)  
 C:Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 17-Apr-1998  
 C:Accession: S74112  
 R:Schnapf, D.; Kemp, G.D.; Smith, V.J.  
 Eur. J. Biochem. 240, 532-539, 1996  
 A:Title: Purification and characterization of a proline-rich antibacterial peptide, with A:Reference number: S74112; MUID:97008941; PMID:8856051  
 A:Accession: S74112  
 A:Molecule type: protein  
 A:Residues: 1-30 <SCH>  
 A:Experimental source: haemocytes  
 C:Keywords: antibacterial

Query Match 16.7%; Score 32; DB 2; Length 30;  
 Best Local Similarity 45.5%; Pred. No. 7.4e+02;  
 Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 5 PYGTLPVGP 15  
 |||||  
 Db 8 PPRPPIGPR 18

## RESULT 12

PC2296  
 v3 domain peptide P4346 - human immunodeficiency virus type 1 (fragment)  
 C:Species: human immunodeficiency virus type 1, HIV-1  
 C:Date: 21-Feb-1995 #sequence\_revision 05-Apr-1995 #text\_change 07-May-1999  
 C:Accession: PC2296  
 R:Sherefa, K.; Soenherborg, A.; Steinberg, J.; Saelberg, M.  
 Biochem. Biophys. Res. Commun. 205, 1658-1664, 1994  
 A:Title: Rapid grouping of HIV-1 infection in subtypes A to E by V3 peptide serotyping a  
 A:Reference number: PC2291; MUID:95110306; PMID:7811250  
 A:Accession: PC2296  
 A:Molecule type: protein  
 A:Residues: 1-35 <SHE>  
 C:Superfamily: type E retrovirus env polyprotein

Query Match 16.7%; Score 32; DB 2; Length 35;  
 Best Local Similarity 46.2%; Pred. No. 8.8e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 8 TLPVGPPEPKT 20  
 :|||:|||||  
 Db 11 SIHMGPRAFYTT 23

## RESULT 13

B41161  
 2XK antigen PEB2 - Campylobacter jejuni (fragment)  
 C:Species: Campylobacter jejuni  
 C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 23-Mar-1993  
 C:Accession: B41161  
 R:Pei, Z.; Blislon III, R.T.; Blaser, M.J.  
 J. Biol. Chem. 266, 16363-16369, 1991  
 A:Title: Identification, purification, and characterization of major antigenic proteins  
 A:Reference number: A41161; MUID:91358413; PMID:1885571  
 A:Accession: B41161  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-35 <PEI>

Query Match 16.7%; Score 32; DB 2; Length 35;  
 Best Local Similarity 47.1%; Pred. No. 8.8e+02;  
 Matches 8; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

OY 2 EVHPYGTLPVGPPEPK 18  
 :|||:|||||  
 Db 1 EILVYG--PGGPAPVPLK 15

## RESULT 14

AD2525  
 hypothetical protein asr7380 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AD2525  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Itiguchi, N.; Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AD2525  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-35 <KUR>

A;Cross-references: GB:BA000020; PIN:BAH7138.1; PID:gl134579; GSPDB:GN00180  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: asr7380  
A;Genome: plasmid

Query Match	16.4%	Score 31.5;	DB 2;	Length 35;
Best Local Similarity	40.7%;	Pred. No. 1e+03;		
Matches 11;	Conservative 1;	Mismatches 2;	Indels 13;	Gaps 2;

QY 11 VGPSPFKTLRVQN-----LGC 28  
|||  
|||:  
Db 6 VGDG---DTLRVRNQGGPITRLGC 28

RESULT 15

H-transporing two-sector ATPase (EC 3.6.3.14) chain G - Chlamydomonas reinhardtii (fr  
N.Alternate names: ATP synthase chain G  
C.Species: Chlamydomonas reinhardtii  
C.Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 03-Jun-2002  
C.Accession: S68394  
R.Fiedler, H.R.; Schmid, R.; Lev, S.; Shavit, N.; Strommann, H.  
A.Title: Isolation of Cf(O)Cf(1) from Chlamydomonas reinhardtii cw15 and the N-terminal.  
A.Reference number: S68398; MUID:96128220; PMID:8543042  
A.Accession: S68394  
A.Molecule type: protein  
A.Residues: 1-19 <PIE>  
A.Experimental source: strain CW15  
C.Genetics:  
A.Genome: nuclear  
C.Keywords: chloroplast; hydrolase

Query Match	16.1%;	Score 31;	DB 2;	Length 19;
Best Local Similarity	40.0%;	Pred. No. 6.2e+02;		
Matches	6;	Conservative	3;	Mismatches 6;
			Indels	0;
			Gaps	0;

```
QY      2 EVHPYGTLPVGPPE 16
          :: : ||| |
Db      4 KIFDNTLPVMAGEE 18
```

Search completed: July 16, 2003, 08:25:26  
Job time : 40 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:15:09 ; Search time 22 Seconds  
(without alignments)  
65,985 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192  
Sequence: 1 YEVAHPYGLPVGPSPFKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 2327

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	34	17.7	28	1	C1QC_RAT
2	32	16.7	30	1	AP65_CARMA
3	31	16.1	34	1	HEMO_CHICK
4	28	14.6	19	1	COOT_SARBU
5	28	14.6	28	1	PA22_MICNI
6	27	14.1	17	1	NIDM_TIRIU
7	27	14.1	22	1	RI18_HALME
8	27	14.1	23	1	RI18_HALVO
9	27	14.1	26	1	RI18_HALNA
10	27	14.1	27	1	PA21_MICNI
11	27	14.1	28	1	PA23_MICNI
12	27	14.1	30	1	PAC2_PSEPO
13	27	14.1	30	1	RI18_HALCU
14	27	14.1	31	1	CY06_VI00D
15	27	14.1	31	1	CY06_VI00D
16	27	14.1	32	1	NEUB_PIG
17	27	14.1	33	1	Y656_TREPA
18	26	13.5	18	1	CRAX_BOVIN
19	26	13.5	30	1	NJSM_PISOC
20	25	13.0	13	1	HMD_METWO
21	25	13.0	19	1	BPT1_BOTUA
22	25	13.0	15	1	MCRA_METTE
23	24.5	12.8	30	1	RBS_PHPYA
24	24.5	12.8	30	1	FIBR_PANIN
25	24.5	12.8	32	1	CA1_ANGUA
26	24	12.5	7	1	UN06_PINPS
27	24	12.5	15	1	UC08_MAIZE
28	24	12.5	18	1	AHD2_TETPY
29	24	12.5	24	1	CAMT_PINPS
30	24	12.5	24	1	RAN_XENLA
31	24	12.5	26	1	CATG_RAT
32	24	12.5	31	1	ER29_BOVIN
33	24	12.5	32	1	PA22_AGKHP

34	24	12.5	35	1	CECB_ANTPE
35	24	12.5	35	1	SERC_PIG
36	23.5	12.2	30	1	PLMS_SQUAC
37	23	12.0	16	1	HBD_GLOPA
38	23	12.0	16	1	LE05_BIOGL
39	23	12.0	27	1	PSAF_SYNYU
40	23	12.0	29	1	PSAF_SYNP6
41	23	12.0	31	1	CIRB_CHAPA
42	23	12.0	31	1	PEPL_CHLVU
43	23	12.0	33	1	PAP1_PAPV
44	23	12.0	33	1	PAP2_PAPV
45	23	12.0	33	1	PAP3_PAPV

## ALIGNMENTS

RESULT 1					
CIQC_RAT		STANDARD	PRT	28 AA	
AC	P31722				
DT	01-JUL-1993 (Rel. 26, Created)				
DT	01-JUL-1993 (Rel. 26, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Complement C1q subcomponent, C chain (Fragment).				
GN	C1Q OR C1QC.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_Taxid=10116;				
RN	[1]				
RP	SEQUENCE.				
RA	MEDLINE:93218657; PubMed:8464426;				
RX	Wing M.G., Seilly D.J., Bridgman R.A.;				
RT	"Rapid isolation and biochemical characterization of rat C1 and C1q."				
RL	Mol. Immunol. 30:433-440(1993).				
CC	-1- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD				
CC	C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE				
CC	COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT				
CC	C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1				
CC	TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE				
CC	FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.				
CC	-1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R				
CC	AND S IN THE MOLAR RATION OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED				
CC	OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE				
CC	A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE				
CC	THE C CHAIN. IN ADDITION TO THE MAJOR A:B AND C:C DIMER BANDS,				
CC	RAT, UNLIKE HUMAN C1Q, CONTAINED MINOR DIMER SPECIES.				
CC	-1- SIMILARITY: CONTAINS 1 COLLAGENOUS DOMAIN.				
DR	InterPro: IPR001073; C1q.				
DR	PROSITE: PS01113; C1Q; PARTIAL.				
KW	Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;				
KW	Repeat.				
FT	DISULFID	4	4		
FT					
FT	MOD_RES	8	8		
FT	MOD_RES	11	11		
FT	MOD_RES	14	14		
FT	MOD_RES	17	17		
FT	NON_TER	28	28		
FT	SEQUENCE	28 AA;	2712 MW;	643F01A24AD590EE CRC64;	
Query Match		17.7%;	Score 34;	DB 1;	Length 28;
Best local similarity		50.0%;	Pred. No. 1.66+02;		
Matches	5;	Conservative	2;	Mismatches	3;
				Indels	0;
				Gaps	0;
QY	25 NLGCVVSGI 34				
DB	1 NAGCYGIPGM 10				
RESULT 2					
AP65_CARMA					

```

ID AP65_CARMA STANDARD; PRT; 30 AA.
AC P82964;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antibacterial 6.5 kDa protein (Fragment).
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
OC Brachyura; Eubrachyura; Portunoidae; Portunidae; Carcinus.
NCBI_TaxID=6759;
RN [1]
RP SEQUENCE, AND FUNCTION.
RX MEDLINE=97008941; PubMed=8856051;
RA Schnapp D., Kemp G.D., Smith V.J.;
RT "Purification and characterization of a proline-rich antibacterial
RT peptide, with sequence similarity to bactericin-7, from the haemocytes
RT of the shore crab, Carcinus maenas."
RL Eur. J. Biochem. 240:532-539 (1996).
CC -1- FUNCTION: STRONG ANTIMICROBIAL ACTIVITY AGAINST P. IMMOBILIS AND
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED MW IS: 6.5 kDa.
CC -1- SIMILARITY: TO BOVINE BACTERICIN 7.
KW Antibiotic.
FT NON TER
SQ SEQUENCE 30 AA; 3305 MW; 6E2C2205934896C4 CRC64;

Query Match 16.7%; Score 32; DB 1; Length 30;
Best Local Similarity 45.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 PYGTLVPVGP 15
DB 8 PPRPPRPPR 18

RESULT 3
HEMO_CHICK
ID HEMO_CHICK STANDARD; PRT; 34 AA.
AC P20057;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemopexin (Fragment).
GN HPX
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN [1]
RP SEQUENCE.
RX MEDLINE=88339942; PubMed=3421961;
RA Wellner D., Cheng K.C., Mueller-Eberhard U.;
RT "N-terminal amino acid sequences of the hemopexins from chicken, rat
RT and rabbit."
RL Biochem. Biophys. Res. Commun. 155:623-625 (1988).
CC -1- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN
CC AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE
CC CIRCULATION.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC PIR: C31514; C31514.
DR InterPro: IPR000585; Hemopexin.
DR PROSITE: PS00024; HEMOPEXIN, PARTIAL.
KW Glycoprotein; Heme; Plasma; Repeat; Transport.
FT NON TER
SQ SEQUENCE 34 AA; 3598 MW; 07A2EAC27386C57 CRC64;

Query Match 16.1%; Score 31; DB 1; Length 34;
Best Local Similarity 55.6%; Pred. No. 5.2e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 4 HPYGLTPVGP 12
DB 15 HPHGAEPPG 23

RESULT 4
COOT_SARBU
ID COOT_SARBU STANDARD; PRT; 19 AA.
AC 009148;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEB-colllostatin (Polliclostatin).
OS Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Oestroidea; Sarcophagidae; Sarcophaga.
NCBI_TaxID=7385;
RN [1]
RP SEQUENCE.
RX MEDLINE=95188911; PubMed=7883009;
RA Byllemans D., Proost P., Samijn B., Borovsky D., Grauwels L.,
RA Huybrechts R., van Damme J., van Beeumen J., de Loof A.;
RT "Neb-colllostatin, a second folliculostatin of the grey fleshfly,
RT Neobellieria bullata."
RL Eur. J. Biochem. 228:45-49 (1995).
CC -1- FUNCTION: HAS AN OSTOSTATIC ACTIVITY. IT INHIBITS YOLK SYNTHESIS SO
CC INHIBITING THE SUBSEQUENT YOLK DEPOSITION IN PREVITELLOGENIC
CC FOLLICLES.
CC -1- SIMILARITY: TO THE NONHELICAL REGIONS OF THE COLLAGEN FAMILY.
CC -1- CAUTION: NEB-COLLOSTATIN MAY BE PROTEOLYTICALLY CLEAVED FROM
CC COLLAGEN IV.
SQ SEQUENCE 19 AA; 1880 MW; 41B6AF5F0CB8251 CRC64;

Query Match 14.6%; Score 28; DB 1; Length 19;
Best Local Similarity 55.6%; Pred. No. 7.4e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 PYGTLVPVGP 13
DB 10 PIGPIVGP 18

RESULT 5
PA22_MICNI
ID PA22_MICNI STANDARD; PRT; 28 AA.
AC P21791;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase A2 isozyme 2 (EC 3.1.1.4) (Phosphatidylcholine
DE 2-acylhydrolase) (Fragment).
OS Micurus nigrocinctus (Central American coral snake) (Gargantilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Micurus.
NCBI_TaxID=8635;
RN [1]
RP SEQUENCE.
RX TISSUE=Venom;
RA Mochica-Morales J., Martin B.M., Zamudio F.Z., Poseani L.D.;
RT "Isolation and characterization of three toxic phospholipases from
RT the venom of the coral snake Micurus nigrocinctus."
RL Toxicon 28:616-617 (1990).
CC -1- FUNCTION: P22 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-
CC ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES. INHIBITS NEUROMUSCULAR
CC TRANSMISSION BY BLOCKING ACETYLCHOLINE RELEASE FROM THE NERVE
CC TERMINI. ACT PRESYNAPTICALLY.
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a fatty acid anion.
CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.

```



## -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.

CC PIR, B35948; B35948.  
 DR HSSP; P15445; 1A3D.  
 DR InterPro: IPR001211; PhospholipaseA2.  
 DR Pfam: PF00068; phosph1p.1  
 DR ProDom: PD000303; PhospholipaseA2; 1.  
 DR PROSITE; PS00118; PA2\_HIS; PARTIAL.  
 DR PROSITE; PS00119; PA2\_ASP; PARTIAL.  
 KW Hydrolase; Lipid degradation; Calcium; Presynaptic neurotoxin; Venom;  
 MultiGene family.  
 FT NON TER 28  
 SO SEQUENCE 28 AA; 3373 MW; 6979852DF2D718BC CRC64;

Query Match 14.6%; Score 28; DB 1; Length 28;  
 Best Local Similarity 55.6%; Pred. No. 1.1e+03;  
 Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 21 LRQNLGCT 29  
 | | | | |  
 Db 18 LFTNYGCT 26

## RESULT 6

ID NU4M TRIRU STANDARD; PRT; 17 AA.  
 AC Q36834;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3) (Fragment).  
 GN ND4.  
 OS Trichophyton rubrum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Onygenales; Arthrodermataceae; mitosporic Arthrodermataceae;  
 OC Trichophyton.  
 NCBI\_TaxID=5551;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IP.1817.89;  
 RA MEDLINE=96132111; PubMed=8593686;  
 RX de Bievre C., Dujon B.;  
 RT "Organisation of the mitochondrial genome of Trichophyton rubrum. DNA  
 RT sequence analysis of the ND4 gene, the ATPase subunit-6 gene, the  
 RT ribosomal RNA small-subunit gene, the ND6 gene, the COXIII gene, the  
 RT ATPase subunit-8 gene and six tRNA genes that correspond respectively  
 RT to the cytosine, lysine, glutamine, asparagine, isoleucine and  
 RT tryptophan isocodons.";  
 RL Curr. Genet. 28:553-559(1995).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL; X88896; CAA61354.1; -  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
 FT NON TER 1  
 SO SEQUENCE 17 AA; 1834 MW; 225A167CE76F84BA CRC64;

Query Match 14.1%; Score 27; DB 1; Length 17;  
 Best Local Similarity 85.7%; Pred. No. 9.1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 29 YVVGSLI 35  
 | | | | |  
 Db 9 YVVGSLI 15

## RESULT 7

ID RL18 HALME STANDARD; PRT; 22 AA.  
 AC P50561;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 50S ribosomal protein L18P (Hmel18) (Fragment).  
 GN RPL18P.  
 OS Halobacterium mediterranei (Haloflex mediterranei).  
 OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;  
 OC Halobacteriaceae; Haloflex.  
 CC NCBI\_TaxID=2252;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=DSM 1411;  
 RA MEDLINE=94229075; PubMed=8174557;  
 RX McDougall J., Wilmann-Liebold B.;  
 RT "Comparative analysis of the protein components from 5S rRNA, protein  
 RT complexes of halophilic archaeobacteria.";  
 RL Eur. J. Biochem. 221:779-785(1994).  
 CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.

DR PIR; P33084; F33084.  
 KW RIBOSOMAL PROTEIN.  
 FT NON TER 22  
 SO SEQUENCE 22 AA; 2773 MW; FE1F607FCAE9876D CRC64;

Query Match 14.1%; Score 27; DB 1; Length 22;  
 Best Local Similarity 44.4%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 14 GPERKTLR 22  
 | | | | |  
 Db 3 GPRYKVPWR 11

## RESULT 8

ID RL18 HALVO STANDARD; PRT; 23 AA.  
 AC P50563;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 50S ribosomal protein L18P (HVOL18) (Fragment).  
 GN RPL18P.  
 OS Halobacterium volcanii (Haloflex volcanii).  
 OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;  
 OC Halobacteriaceae; Haloflex.  
 CC NCBI\_TaxID=2246;  
 RN [1]  
 RP SEQUENCE.

RC STRAIN=DSM 3757;  
 MEDLINE=94229075; PubMed=8174557;  
 RX McDougall J., Wilmann-Liebold B.;  
 RT "Comparative analysis of the protein components from 5S rRNA, protein  
 RT complexes of halophilic archaeobacteria.";  
 RL Eur. J. Biochem. 221:779-785(1994).  
 CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.

DR PIR; I33084; I33084.  
 FT NON TER 23  
 SO SEQUENCE 23 AA; 2929 MW; 246E1F607FCAE987 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 23;  
 Best Local Similarity 44.4%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 14 GPERKTLR 22  
 | | | | |  
 Db 3 GPRYKVPWR 11

## RESULT 9

RL18\_HALHA

```

ID RL18 HALHA STANDARD; PRT; 26 AA.
AC PS0560;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L18p (HHA18) (Fragment).
GN RPL18P.
OS Halobacterium halobium.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN (1)
RP SEQUENCE.
RC STRAIN=DSM 670;
RX MEDLINE=94229075; PubMed=8174557;
RA McDougall J., Wittmann-Liebold B.;
RT "Comparative analysis of the protein components from 5S rRNA-protein
RL complexes of halophilic archaeobacteria.";
RL Eur. J. Biochem. 221:779-785(1994).
CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
DR PIR: B33084; B33084.
KW Ribosomal protein.
FT NON_TER 26
SQ SEQUENCE 26 AA; 3269 MW; D569CB746E1F607F CRC64;

Query Match 14.1%; Score 27; DB 1; Length 26;
Best Local Similarity 44.4%; Pred. No. 1.4e+03;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 14 GPEFKTLR 22
DB 3 GPRYKVPWR 11

RESULT 10
PA23 MICNI STANDARD; PRT; 27 AA.
ID PA23 MICNI
AC P21790;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase A2 isozyme 1 (EC 3.1.1.4) (Phosphatidylcholine
2-acetylhydrolase) (Fragment).
OS Microtus nigricinctus (Central American coral snake) (Gargantilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Micrurus.
OX NCBI_TaxID=8635;
RN (1)
RP SEQUENCE.
RC TISSUE=Venom;
RA Mochica-Morales J., Martin B.M., Zamudio F.Z., Posanti L.D.;
RT "Isolation and characterization of three toxic phospholipases from
RT the venom of the coral snake Micrurus nigricinctus.";
RT Toxicon 28:616-617(1990).
CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-
ACYL GROUPS IN 3-SN-PHOSPHOLIPIDS. INHIBITS NEUROMUSCULAR
TRANSMISSION BY BLOCKING ACETYLCHOLINE RELEASE FROM THE NERVE
TERMINI. ACT PRESYNAPTICALLY.
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
acylglycerophosphocholine + a fatty acid anion.
CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
DR PIR: A35948; A35948.
DR HSSP: P00598; 1POA.
DR InterPro: IPR001211; PhospholipaseA2.
DR Pfam: PF00068; Phoslip; 1.
DR PROSITE: PS00118; PA2_HIS; PARTIAL.
DR PROSITE: PS00119; PA2_ASP; PARTIAL.
KW Hydrolyase, lipid degradation; Calcium; Presynaptic neurotoxin; Venom;
KW Multigene family.
FT NON_TER 27

```

```

SQ SEQUENCE 27 AA; 3314 MW; 38637ECA600F49A0 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 27;
Best Local Similarity 33.3%; Pred. No. 1.5e+03;
Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 18 KTLRFVQNIQCY 29
DB 15 RSVLEFMFGYCY 26

RESULT 11
PA23 MICNI STANDARD; PRT; 28 AA.
ID PA23 MICNI
AC P21792;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase A2 isozyme 3 (EC 3.1.1.4) (Phosphatidylcholine
2-acetylhydrolase) (Fragment).
OS Microtus nigricinctus (Central American coral snake) (Gargantilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Elapinae; Micrurus.
OX NCBI_TaxID=8635;
RN (1)
RP SEQUENCE.
RC TISSUE=Venom;
RA Mochica-Morales J., Martin B.M., Zamudio F.Z., Posanti L.D.;
RT "Isolation and characterization of three toxic phospholipases from
RT the venom of the coral snake Micrurus nigricinctus.";
RT Toxicon 28:616-617(1990).
CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-
ACYL GROUPS IN 3-SN-PHOSPHOLIPIDS. INHIBITS NEUROMUSCULAR
TRANSMISSION BY BLOCKING ACETYLCHOLINE RELEASE FROM THE NERVE
TERMINI. ACT PRESYNAPTICALLY.
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
acylglycerophosphocholine + a fatty acid anion.
CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
DR PIR: C35948; C35948.
DR HSSP: P15445; 1A3D.
DR InterPro: IPR001211; PhospholipaseA2.
DR Pfam: PF00068; Phoslip; 1.
DR PROSITE: PS00118; PA2_HIS; PARTIAL.
DR PROSITE: PS00119; PA2_ASP; PARTIAL.
KW Hydrolyase, lipid degradation; Calcium; Presynaptic neurotoxin; Venom;
KW Multigene family.
FT NON_TER 28
SQ SEQUENCE 28 AA; 3394 MW; 315FB012F69098B1 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 28;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 25 NLGCV 29
DB 22 NAGCV 26

RESULT 12
PA2C_PSEPO STANDARD; PRT; 28 AA.
ID PA2C_PSEPO
AC P20260;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase A2 (EC 3.1.1.4) (Pseudexin C chain) (Phosphatidylcholine
2-acetylhydrolase) (Fragment).
OS Pseudexin porphyriacus (Red-bellied black snake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Acanthophinae; Pseudechis.
OX NCBI_TaxID=8671;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=8938835; PubMed=2675391;
RA Schmidt J.J., Middlebrook J.L.;
RT "Purification, sequencing and characterization of pseudexin
RT phospholipases A2 from Pseudechis porphyriacus (Australian
RT red-bellied black snake).";
RL Toxicon 27:805-818(1989).
CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE
CC -2- ACYL GROUPS IN 3-SN-PHOSPHOLIPIDS.
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a fatty acid anion.
CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
DR PIR: C32416; C32416.
DR HSSP: P00592; 2PH1.
DR InterPro: IPR001211; PhospholipaseA2.
DR Pfam: PF000068; phospho_1.
DR PRODOM: PD000303; PhospholipaseA2; 1.
DR PROSITE: PS00118; PA2_HIS; PARTIAL.
DR PROSITE: PS00119; PA2_ASP; PARTIAL.
KW Hydrolase; Lipid degradation; Calcium; Multigene family; Venom.
FT NON TER 28
SQ SEQUENCE 28 AA; 3210 MW; 5089A7E85CAE0D5 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 28;
Best Local Similarity 38.9%; Pred. No. 1.5e+03;
Matches 7; Conservative 2; Mismatches 5; Indels 4; Gaps 1;

OY 12 GPGPEFKTIRQNIGCY 29
DB 15 GSRPLFYR---DYGCY 28

RESULT 13
RL18 HALCU
ID RL18 HALCU STANDARD; PRT; 30 AA.
AC P05970;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L18P (HCU18) (HL13) (Fragment).
GN Pp18P.
OS Halobacterium cutirubrum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE.
RX MEDLINE=79045279; PubMed=152199;
RX Smith N., Matheson A.T., Yaguchi M., Willick G., Nazat R.N.;
RT "The 5-S RNA-protein complex from an extreme halophile.
RT Halobacterium cutirubrum. Purification and characterization.";
RL Eur. J. Biochem. 89:501-509(1978).
CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
DR PIR: S07217; S07217.
KW Ribosomal protein.
FT NON TER 30
SQ SEQUENCE 30 AA; 3624 MW; 3A50079B1569CB74 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 30;
Best Local Similarity 44.4%; Pred. No. 1.7e+03;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 14 GPEFKTIR 22
DB 3 GPRYKVPKR 11

```

```

RESULT 14
ID CYOB VIOOD STANDARD; PRT; 31 AA.
AC P58439;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cycloviolacin O6.
OS Viola odorata (Sweet Violet).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosida I; Malpighiales; Violaceae; Viola.
OX NCBI_TaxID=97441;
RN [1]
RP SEQUENCE.
RX MEDLINE=20069951; PubMed=10600388;
RX Craik D.J., Daly N.L., Bond T., Waine C.;
RT "Plant cyclotides: a unique family of cyclic and knotted proteins that
RT defines the cyclic cysteine knot structural motif.";
RL J. Mol. Biol. 294:1327-1336(1999).
CC -1- FUNCTION: Probably participates in a plant defense mechanism.
CC -1- PTM: This is a cyclic peptide.
CC -1- CAUTION: This peptide is cyclic, its sequence was chosen to start
CC at the position shown below by similarity to Oak1 (Kallata B1)
CC whose DNA sequence is known.
DR DISULFID 5 21
DR DISULFID 9 23
DR DISULFID 14 28
SQ SEQUENCE 31 AA; 3236 MW; 30D6A52F72C6E68 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 31;
Best Local Similarity 83.3%; Pred. No. 1.7e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 7 GTLPVG 12
DB 1 GTLPVG 6

RESULT 15
ID CYOB VIOOD STANDARD; PRT; 31 AA.
AC P58439;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cycloviolacin O11.
OS Viola odorata (Sweet Violet).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosida I; Malpighiales; Violaceae; Viola.
OX NCBI_TaxID=97441;
RN [1]
RP SEQUENCE.
RX MEDLINE=20069951; PubMed=10600388;
RX Craik D.J., Daly N.L., Bond T., Waine C.;
RT "Plant cyclotides: a unique family of cyclic and knotted proteins that
RT defines the cyclic cysteine knot structural motif.";
RL J. Mol. Biol. 294:1327-1336(1999).
CC -1- FUNCTION: Probably participates in a plant defense mechanism.
CC -1- PTM: This is a cyclic peptide.
CC -1- CAUTION: This peptide is cyclic, its sequence was chosen to start
CC at the position shown below by similarity to Oak1 (Kallata B1)
CC whose DNA sequence is known.
DR DISULFID 5 21
DR DISULFID 9 23
DR DISULFID 14 28
SQ SEQUENCE 31 AA; 3236 MW; 30D6A52F72C6E68 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 31;
Best Local Similarity 83.3%; Pred. No. 1.7e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Best Local Similarity 83.3%; Pred. No. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 GTLPVG 12  
|||  
|  
Db 1 GTLPVG 6

Search completed: July 16, 2003, 08:23:15  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:20:19 ; Search time 78 Seconds  
(without alignments)  
92.457 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192  
Sequence: 1 YEHPHYGTLPVGPGEFRTLRVQNLGCVWSGLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 20244

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	24.0	34	15	O91354 human immun
2	46	24.0	34	15	O91355 human immun
3	46	24.0	34	15	O91357 human immun
4	46	24.0	34	15	O78489 human immun
5	45	23.4	34	15	O78487 human immun
6	45	23.4	35	15	O91436 human immun
7	45	23.4	35	15	O91437 human immun
8	45	23.4	35	15	O90519 human immun
9	45	23.4	35	15	O76251 human immun
10	44	22.9	35	15	O91434 human immun
11	44	22.9	35	15	O91454 human immun
12	44	22.9	35	15	O80499 human immun
13	44	22.9	35	15	O10869 human immun
14	44	22.9	35	15	O76250 human immun
15	44	22.9	35	15	O76296 human immun
16	43	22.4	34	15	O91356 human immun

17	43	22.4	34	15	O70742 human immun
18	43	22.4	34	15	O70743 human immun
19	43	22.4	34	15	O70747 human immun
20	43	22.4	35	15	O91349 human immun
21	43	22.4	35	15	O91350 human immun
22	43	22.4	35	15	O91351 human immun
23	43	22.4	35	15	O91430 human immun
24	43	22.4	35	15	O91433 human immun
25	43	22.4	35	15	O91435 human immun
26	43	22.4	35	15	O91438 human immun
27	43	22.4	35	15	O91439 human immun
28	43	22.4	35	15	O91440 human immun
29	43	22.4	35	15	O91446 human immun
30	43	22.4	35	15	O90503 human immun
31	43	22.4	35	15	O9WCW5 human immun
32	43	22.4	35	15	O9CCU4 human immun
33	43	22.4	35	15	O80465 human immun
34	43	22.4	35	15	O80466 human immun
35	43	22.4	35	15	O80467 human immun
36	43	22.4	35	15	O78043 human immun
37	43	22.4	35	15	O78055 human immun
38	43	22.4	35	15	O78206 human immun
39	43	22.4	35	15	O78212 human immun
40	43	22.4	35	15	O78215 human immun
41	43	22.4	35	15	O9WCW6 human immun
42	43	22.4	35	15	O70232 human immun
43	43	22.4	35	15	O79420 human immun
44	43	22.4	35	15	O76295 human immun
45	43	22.4	35	15	O77931 human immun

## ALIGNMENTS

RESULT 1					
ID	O91354	PRELIMINARY;	PRT;	34 AA.	
AC	O91354;				
DT	01-NOV-1998 (TREMBLrel. 08, Created)				
DT	01-NOV-1998 (TREMBLrel. 09, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Envelope glycoprotein (Fragment).				
GN	ENV.				
OS	Human immunodeficiency virus type 1.				
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.				
OX	NCBI_TaxID=11676;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=TK-22;				
RX	MEDLINE=98090117; PubMed=9430252;				
RA	Ida S., Gatanaga H., Shioda T., Nagai Y., Kobayashi N., Shinada K.,				
RA	Kimura S., Iwamoto A., Oka S.,				
RT	"HIV type 1 V3 variation dynamics in vivo: long-term persistence of				
RT	non-synonymous-inducing genotypes and transient presence of syncytium-				
RT	inducing genotypes during the course of progressive AIDS.";				
RL	AIDS Res. Hum. Retroviruses 13:1597-1609(1997).				
DR	EMBL; AB005326; BAA33244.1;				
KW	AIDS; Coat protein; Glycoprotein.				
FT	NON_TER	1			
FT	NON_TER	34			
FT	NON_TER	1			
SQ	SEQUENCE	34 AA;	3785 MW;	A31972064FAD828F CRC64;	
Query Match					
Best Local Similarity 24.0%; Score 46; DB 15; Length 34;					
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;					
OY 8 TLVPGPGEFRTLRVQNL 26					
DB 11 SLIPGPRAFYTNIGNI 29					
RESULT 2					
O91355					

ID O91355 PRELIMINARY; PRT; 34 AA.  
AC O91355;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK-22;  
RX MEDLINE=98090117; PubMed=9430252;  
RA Ida S., Gatanga H., Shida T., Nagai Y., Kobayashi N., Shimada K.,  
RA Kimura S., Iwamoto A., Oka S.;  
RT "HIV type 1 V3 variation dynamics in vivo: long-term persistence of  
RT non-synctium-inducing genotypes and transient presence of cynctium-  
RT inducing genotypes during the course of progressive AIDS.";  
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).  
DR EMBL; AB05327; BAA33245.1; -;  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1 1  
FT NON\_TER 34 34  
SQ SEQUENCE 34 AA; 3785 MW; A31972064FAD828F CRC64;  
  
Query Match 24.0%; Score 46; DB 15; Length 34;  
Best Local Similarity 42.1%; Pred. No. 15;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 8 TLVPGPPEFKTTLRVQNL 26  
DB 11 SIPIGPGRAFTYTTNIGNI 29

RESULT 3  
ID O91357 PRELIMINARY; PRT; 34 AA.  
AC O91357;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK-22;  
RX MEDLINE=98090117; PubMed=9430252;  
RA Ida S., Gatanga H., Shida T., Nagai Y., Kobayashi N., Shimada K.,  
RA Kimura S., Iwamoto A., Oka S.;  
RT "HIV type 1 V3 variation dynamics in vivo: long-term persistence of  
RT non-synctium-inducing genotypes and transient presence of cynctium-  
RT inducing genotypes during the course of progressive AIDS.";  
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).  
DR EMBL; AB05329; BAA33247.1; -;  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1 1  
FT NON\_TER 34 34  
SQ SEQUENCE 34 AA; 3811 MW; A303C2064FAD828F CRC64;

Query Match 24.0%; Score 46; DB 15; Length 34;  
Best Local Similarity 42.1%; Pred. No. 15;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
  
QY 8 TLVPGPPEFKTTLRVQNL 26  
DB 11 SIPIGPGRAFTYTTNIGNI 29

RESULT 4

Q78489  
ID Q78489 PRELIMINARY; PRT; 34 AA.  
AC Q78489;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Viral sample FLO5R3C (Florida local control 01), partial env cds, V3  
DE region (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Zhang L.O., Leigh-Brown A.J.;  
RL Submitted (Apr-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92271245; PubMed=1589796;  
RA Ou C.-Y., Ciesielski C.A., Myers G., Banea C.I., Luo C.C.,  
RA Korber B.T.M., Mullins J.I., Schochetman G., Berkelman R.L.,  
RA Economou A.N., Witte J.J., Furman L.J., Satten G.A., Curran J.W.,  
RA Jaffe H.W.;  
RT "Molecular Epidemiology of HIV Transmission in a Dental Practice";  
RL Science 256:1165-1171(1992).  
DR EMBL; M92128; AAA44576.1; -;  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1 1  
FT NON\_TER 34 34  
SQ SEQUENCE 34 AA; 3786 MW; A93F03A44FAD828F CRC64;  
  
Query Match 24.0%; Score 46; DB 15; Length 34;  
Best Local Similarity 42.1%; Pred. No. 15;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 8 TLVPGPPEFKTTLRVQNL 26  
DB 11 SIPIGPGRAFTYTTNIGNI 29

RESULT 5  
ID Q78487 PRELIMINARY; PRT; 34 AA.  
AC Q78487;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Viral sample FLO5R3A (Florida local control 01), partial env cds, V3  
DE region (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Zhang L.O., Leigh-Brown A.J.;  
RL Submitted (Apr-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92271245; PubMed=1589796;  
RA Ou C.-Y., Ciesielski C.A., Myers G., Banea C.I., Luo C.C.,  
RA Korber B.T.M., Mullins J.I., Schochetman G., Berkelman R.L.,  
RA Economou A.N., Witte J.J., Furman L.J., Satten G.A., Curran J.W.,  
RA Jaffe H.W.;  
RT "Molecular Epidemiology of HIV Transmission in a Dental Practice";  
RL Science 256:1165-1171(1992).  
DR EMBL; M92127; AAA44575.1; -;  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1 1  
FT NON\_TER 34 34  
SQ SEQUENCE 34 AA; 3818 MW; A93F02E148C0428F CRC64;

Query Match 23.4%; Score 45; DB 15; Length 34;

Best Local Similarity 42.1%; Pred. No. 21;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
Qy 8 TLPVGPGEFPTLRVONT 26  
:::|||||:::  
Db 11 SIPMGPGAFYTTTROI 29

## RESULT 6

O91436 PRELIMINARY; PRT; 35 AA.  
AC O91436.  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK-2;  
RX MEDLINE=96090117; PubMed=9430252;  
RA Ida S., Gatnaga H., Shioda T., Nagai Y., Kobayashi N., Shimada K.,  
RA Kimura S., Iwamoto A., Oka S.;  
RT "HIV type 1 VA variation dynamics in vivo: long-term persistence of  
RT non-synctium-inducing genotypes and transient presence of cynctium-  
inducing genotypes during the course of progressive AIDS.";  
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).  
DR EMBL; AB005421; BAA3338.1; -  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1  
FT 35  
SQ SEQUENCE 35 AA; 3956 MW; 9F5A5B3A688BF77E CRC64;

Query Match 23.4%; Score 45; DB 15; Length 35;  
Best Local Similarity 43.8%; Pred. No. 22;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 8 TLPVGPGEFPTLRV 23  
:::|||||:::  
Db 11 SIPMGPGAFYTTROI 26

## RESULT 7

O91437 PRELIMINARY; PRT; 35 AA.  
AC O91437.  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK-2;  
RX MEDLINE=96090117; PubMed=9430252;  
RA Ida S., Gatnaga H., Shioda T., Nagai Y., Kobayashi N., Shimada K.,  
RA Kimura S., Iwamoto A., Oka S.;  
RT "HIV type 1 VA variation dynamics in vivo: long-term persistence of  
RT non-synctium-inducing genotypes and transient presence of cynctium-  
inducing genotypes during the course of progressive AIDS.";  
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).  
DR EMBL; AB005422; BAA3339.1; -  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1  
FT 35  
SQ SEQUENCE 35 AA; 3956 MW; 9F5A5B3A688BF77E CRC64;

Query Match 23.4%; Score 45; DB 15; Length 35;  
Best Local Similarity 43.8%; Pred. No. 22;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
Qy 8 TLPVGPGEFPTLRV 23  
:::|||||:::  
Db 11 SIPMGPGAFYTTROI 26

## RESULT 8

O90519 PRELIMINARY; PRT; 35 AA.  
AC O90519.  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=745;  
RX MEDLINE=96090117; PubMed=9430252;  
RA Ida S., Gatnaga H., Shioda T., Nagai Y., Kobayashi N., Shimada K.,  
RA Kimura S., Iwamoto A., Oka S.;  
RT "HIV type 1 VA variation dynamics in vivo: long-term persistence of  
RT non-synctium-inducing genotypes and transient presence of cynctium-  
inducing genotypes during the course of progressive AIDS.";  
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).  
DR EMBL; AB005421; BAA3338.1; -  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1  
FT 35  
SQ SEQUENCE 35 AA; 3956 MW; 9F5A5B3A688BF77E CRC64;

Query Match 23.4%; Score 45; DB 15; Length 35;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 8 TLPVGPGEFPTLRV 23  
:::|||||:::  
Db 11 SIPMGPGAFYTTROI 26

## RESULT 9

O76251 PRELIMINARY; PRT; 35 AA.  
AC O76251.  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE V3 loop (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M.G.;  
RX MEDLINE=96090117; PubMed=9430252;  
RA Ida S., Gatnaga H., Shioda T., Nagai Y., Kobayashi N., Shimada K.,  
RA Kimura S., Iwamoto A., Oka S.;  
RT "HIV type 1 VA variation dynamics in vivo: long-term persistence of  
RT non-synctium-inducing genotypes and transient presence of cynctium-  
inducing genotypes during the course of progressive AIDS.";  
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).  
DR EMBL; AB005422; BAA3339.1; -  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1  
FT 35  
SQ SEQUENCE 35 AA; 3956 MW; 9F5A5B3A688BF77E CRC64;

Query Match 23.4%; Score 45; DB 15; Length 35;  
Best Local Similarity 43.8%; Pred. No. 22;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 8 TLVGGPPEFKTLRV 23  
 Db 11 SIPIGGRAFYTTROI 26

## RESULT 10

091434 PRELIMINARY; PRT; 35 AA.  
 AC 091434;  
 DT 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Envelope glycoprotein (Fragment).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OK NCBI\_Taxid=11676;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TK-2;  
 RX MEDLINE=98090117; PubMed=9430252;  
 RA Ida S., Gatanga H., Shioda T., Nagai Y., Kobayashi N., Shimada K., Kimura S., Iwamoto A., Oka S.,  
 RT "HIV type 1 V3 variation dynamics in vivo: long-term persistence of non-synctium-inducing genotypes and transient presence of cynctium-inducing genotypes during the course of progressive AIDS.";  
 RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).  
 DR EMBL; AB005419; BAA3336.1; -  
 KM AIDS; Coat protein; Glycoprotein.  
 FT NON\_TER 1  
 SQ SEQUENCE 35 AA; 3831 MW; 9F5ASB2A698BE26B CRC64;  
 OY 8 TLVGGPPEFKTT 20  
 Db 11 SIPIGGRAFYTT 23

## RESULT 11

091454 PRELIMINARY; PRT; 35 AA.  
 AC 091454;  
 DT 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Envelope glycoprotein (Fragment).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OK NCBI\_Taxid=11676;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TK-2;  
 RX MEDLINE=98090117; PubMed=9430252;  
 RA Ida S., Gatanga H., Shioda T., Nagai Y., Kobayashi N., Shimada K., Kimura S., Iwamoto A., Oka S.,  
 RT "HIV type 1 V3 variation dynamics in vivo: long-term persistence of non-synctium-inducing genotypes and transient presence of cynctium-inducing genotypes during the course of progressive AIDS.";  
 RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).  
 DR EMBL; AB005440; BAA33356.1; -  
 KM AIDS; Coat protein; Glycoprotein.  
 FT NON\_TER 1  
 SQ SEQUENCE 35 AA; 4026 MW; 9E4AA4A3A68BF77E CRC64;

Query Match 22.9%; Score 44; DB 15; Length 35;  
 Best Local Similarity 46.7%; Pred. No. 30;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 9 LPVGGPPEFKTLRV 23  
 Db 12 IPIGGRAFYTTROI 26

## RESULT 12

080499 PRELIMINARY; PRT; 35 AA.  
 AC 080499;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Envelope protein (Fragment).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OK NCBI\_Taxid=11676;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90364416; PubMed=2392685;  
 RA Larosa G.J., Davide J.P., Weinhold K., Waterbury J.A., Profy A.T., Lewis J.A., Langlois A.J., Dreesman G.R., Boswell R.N., Shadduck P., Holley L.H., Karplus M., Bolognesi D.P., Matthews T.J., Emini E.A., Putney S.D.;  
 RT "Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant";  
 RL Science 249:932-935(1990).  
 RN 12  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9118022; PubMed=1990444;  
 RA Larosa G.J., Davide J.P., Weinhold K., Waterbury J.A., Profy A.T., Lewis J.A., Langlois A.J., Dreesman G.R., Boswell R.N., Shadduck P., Holley L.H., Karplus M., Bolognesi D.P., Matthews T.J., Emini E.A., Putney S.D.;  
 RT "Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant: Corrections and clarifications.";  
 RL Science 251:811-811(1991).  
 RN 13  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91361090; PubMed=1887238;  
 RA Larosa G.J., Weinhold K., Profy A.T., Langlois A.J., Dreesman G.R., Boswell R.N., Shadduck P., Bolognesi D.P., Matthews T.J., Emini E.A., Putney S.D.;  
 RT "Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant: Further clarifications.";  
 RL Science 253:1146-1146(1991).  
 DR EMBL; M61532; AAA45222.1; -  
 KM AIDS; Coat protein; Glycoprotein.  
 FT NON\_TER 1  
 SQ SEQUENCE 35 AA; 4022 MW; 0780966F52E750D0 CRC64;

Query Match 22.9%; Score 44; DB 15; Length 35;  
 Best Local Similarity 34.6%; Pred. No. 30;  
 Matches 9; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

OY 3 VHPYGT---LPVGGPPEFKTLRVQ 24  
 Db 2 IPIGGRAFYTTROI 27

## RESULT 13

010869 PRELIMINARY; PRT; 35 AA.  
 AC 010869;  
 DT 01-JUL-1997 (TRENBLrel. 04, Created)  
 DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Envelope glycoprotein (Fragment).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OK NCBI\_Taxid=11676;



```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LIFI;
RX MEDLINE=97186421; PubMed=9115816;
RA Mang B., Ge Y.C., Palasanthiran P., Ziegler J., Bolton W., Xiang S.H.,
RA Dwyer D.E., Cunningham A.L., Saksena N.K.;
RT "HIV type 1 V3 loop sequences derived from peripheral blood of
RT transmitting mothers, their infants, and nontransmitting mothers
RT differ in their crown octapeptide motifs."
RL AIDS Res. Hum. Retroviruses 13:275-279 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LIFI;
RA Blouin J.C.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U66657; AAB58149.1; -.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT SEQUENCE 35 AA; 3718 MW; 9F53DA6E98B8B8C CRC64;
SQ

Query Match 22.9%; Score 44; DB 15; Length 35;
Best Local Similarity 36.0%; Pred. No. 30;
Matches 9; Conservative 5; Mismatches 7; Indels 4; Gaps 1;

Qy 8 TLPGPGPEFKTTLR---VQNLGC 28
Db 11 SIPIGPGRAFTTGEIIGDIRQAGC 35

RESULT 14
Q76250 PRELIMINARY; PRT; 35 AA.
AC Q76250;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
RT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE V3 loop (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Schreiber M.G.;
RT "Loss of V3 domain specific antibodies to selected HIV-1 variants over
RT time is associated with viral escape and selection in vivo."
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z50844; CAA90697.1; -.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT SEQUENCE 35 AA; 3898 MW; 8A1B0F25558BF77E CRC64;
SQ

Query Match 22.9%; Score 44; DB 15; Length 35;
Best Local Similarity 46.7%; Pred. No. 30;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 9 LPVGPPEFKTLRV 23
Db 12 IPIGPGRAFTTQT 26

RESULT 15
Q76296 PRELIMINARY; PRT; 35 AA.
AC Q76296;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
RT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE V3 loop (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.

```

```

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94246736; PubMed=8189527;
RA Schreiber M.G., Petersen H., Wachsmuth C., Muller H., Hufert F.T.,
RA Schmitz H.;
RT "Antibodies of symptomatic human immunodeficiency virus type 1-
RT infected individuals are directed to the V3 domain of noninfectious
RT and not of infectious virions present in autologous serum."
RL J. Virol. 68:3908-3916 (1994).
DR EMBL; Z15163; CAA78862.1; -.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT SEQUENCE 35 AA; 3775 MW; 9F5B905B698BF77E CRC64;
SQ

Query Match 22.9%; Score 44; DB 15; Length 35;
Best Local Similarity 53.8%; Pred. No. 30;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 8 TLPGPGPEFKTT 20
Db 11 SIPIGPGSAFTT 23

Search completed: July 16, 2003, 08:24:40
Job time : 79 secs

```

**THIS PAGE BLANK (USPTO)**





Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

OY 8 TLPVGPGEFKT-----LRVQNLG 27  
 ::||| |  
 :||| |  
 Db 5 SIPIGGRAFYTTGYKYKVKIEPLG 30

RESULT 4  
 AAY39728  
 ID AAY39728 standard; peptide; 33 AA.  
 XX  
 AC AAY39728;  
 XX  
 DT 26-NOV-1999 (first entry)  
 XX  
 DE HIV1 chimeric peptide CLTB-159.  
 XX  
 KW HIV, vaccine; immunogenic composition; T cell epitope; B cell epitope;  
 KM infection; antibody; antiviral.  
 XX  
 OS Human immunodeficiency virus type 1.  
 XX  
 PN US5951986-A.  
 PD 14-SEP-1999.  
 XX  
 PF 06-JUN-1995; 95US-0467881.  
 XX  
 PR 09-JUN-1994; 94US-0257528.  
 PR 09-JUN-1993; 93US-0073378.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Klein MH, Chong P, Sia CDY;  
 XX  
 DR WPI; 1999-550482/46.  
 XX  
 PT Immunogenic composition containing synthetic fusion polypeptides  
 PT containing both the T and B cell epitopes of the human immunodeficiency  
 PT virus, useful antigens in producing vaccines -  
 XX  
 PS Example 1; Column 25-26; 43pp; English.  
 XX  
 CC This sequence represents a fragment of a HIV1 protein, and can be used in  
 CC the immunogenic composition of the invention. The composition comprises a  
 CC synthetic fusion polypeptide which includes a sequence encoding 1 or more  
 CC T cell epitopes and a sequence encoding 1 or more B cell epitopes and a  
 CC carrier. Both the T cell and B cell epitopes are derived from HIV  
 CC proteins. The compositions are useful as vaccines against HIV infection.  
 CC The composition induces HIV-1-specific polyclonal antibodies that are  
 CC opsonising and antiviral. The peptide components may be selected to  
 CC induce a response against different viral isolates and in subjects who  
 CC recognise different T cell epitopes.  
 XX  
 SQ Sequence 33 AA;

Query Match 24.0%; Score 46; DB 20; Length 33;  
 Best Local Similarity 34.6%; Pred. No. 30;  
 Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

OY 8 TLPVGPGEFKT-----LRVQNLG 27  
 ::||| |  
 :||| |  
 Db 5 SIPIGGRAFYTTGYKYKVKIEPLG 30

RESULT 5  
 AAW98900  
 ID AAW98900 standard; peptide; 33 AA.  
 XX  
 AC AAW98900;  
 XX  
 DT 05-MAY-1999 (first entry)  
 XX

DE HIV-1 vaccine synthetic peptide SEQ ID NO:84.  
 XX  
 KM HIV-1; human immunodeficiency virus; vaccine; T-cell epitope;  
 KM gag protein; B-cell epitope; gp41 protein; chimeric; infection.  
 XX  
 OS Synthetic.  
 OS Human immunodeficiency virus type 1.  
 XX  
 PN US5876731-A.  
 PD 02-MAR-1999.  
 XX  
 PF 05-JUN-1995; 95US-0462507.  
 XX  
 PR 09-JUN-1994; 94US-0257528.  
 PR 09-JUN-1993; 93US-0073378.  
 PR 05-JUN-1995; 95US-0462507.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Chong P, Klein MH, Sia CDY;  
 XX  
 DR WPI; 1999-189590/16.  
 XX  
 PT Synthetic chimeric HIV polypeptides - comprising gag protein T-cell  
 PT epitope linked to gp41 B-cell epitope  
 XX  
 PS Example 1; Column 63-64; 41pp; English.  
 CC The present invention describes a synthetic peptide comprising an amino  
 CC acid sequence containing a T-cell epitope of an HIV gag protein linked  
 CC at its C terminus to an amino acid sequence containing a B-cell epitope  
 CC of an HIV gp41 protein and containing the amino acid sequence:  
 CC X1LKDWX2; where X1 = F, A, G or Q, and X2 = A or T, or an amino acid  
 CC sequence capable of eliciting an HIV-specific antiserum and recognizing  
 CC the sequence X1LKDWX2. The synthetic peptide is useful in vaccines  
 CC against HIV infection and in diagnostic applications. AAW98892 to  
 CC AAW98906, and AAW98999 to AAW99899 represent synthetic peptides from the  
 CC present invention.  
 XX  
 SQ Sequence 33 AA;

Query Match 24.0%; Score 46; DB 20; Length 33;  
 Best Local Similarity 34.6%; Pred. No. 30;  
 Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

OY 8 TLPVGPGEFKT-----LRVQNLG 27  
 ::||| |  
 :||| |  
 Db 5 SIPIGGRAFYTTGYKYKVKIEPLG 30

RESULT 6  
 AAR77763  
 ID AAR77763 standard; peptide; 35 AA.  
 XX  
 AC AAR77763;  
 XX  
 DT 08-SEP-1995 (first entry)  
 XX  
 DE PRI/T5 chimera lysine branched peptide.  
 XX  
 KM T-cell; epitope; HIV-1; core protein; p24E; B-cell; antigen;  
 KM gp160; gag; pol; vaccine; multimeric peptide; AIDS; 3D organisation.  
 XX  
 OS Human immunodeficiency virus type 1.  
 XX  
 PH Key  
 PH 1..18 Location/Qualifiers  
 FT Peptide  
 FT /label= PRI  
 FT /note= "B cell epitope"  
 FT 19..33  
 FT /label= T5  
 FT /note= "T cell epitope"

```
FT Modified-site 34 /note= "Modified at the epsilon N by the sequence given
FT /note= "in AAR68713, which represents the same sequence
FT as bases 1-33 of this sequence"
FT Modified-site 35 /note= "Modified at the epsilon N with a Lys residue
FT which is substituted on the alpha and epsilon
FT N with the sequence given in AAR68713"
XX
XX PN WO9429339-A.
XX
XX PD 22-DEC-1994.
XX
XX PF 08-JUN-1994; 94WO-CA00317.
XX
XX PR 09-JUN-1993; 93US-0073378.
XX
XX PA (CONN-) CONNAUGHT LAB LTD.
XX
XX PI Chong P, Klein MH, Sia CDY;
XX
XX DR WPI; 1995-036400/05.
XX
XX PT Novel tandem synthetic HIV-1 peptide(s) - comprising T-cell
XX epitope of gag protein linked to B-cell epitope of V3 loop
XX protein of an HIV-1 isolate
XX
XX PS Claim 4; Page 46; 69pp; English.
XX
XX CC This sequence represents a multimeric lysine branched peptide which
XX contains four copies a peptide monomer comprising the T-cell epitope
XX derived from the HIV-1 core protein, T5, linked to the B-cell epitope,
XX PRI which represents a consensus sequence derived from New York and
XX Amsterdam HIV-1 isolates. The entire multimer contains four copies of
XX the same sequence linked through Lys residues. The B-cell epitope may
XX be derived from HIV-1 proteins esp. gp160, gag and pol proteins. These
XX chimeric peptides may then be used in the production of HIV-1 vaccines.
XX The linear and multimeric peptides may be used for the treatment of AIDS
XX by acting to displace the binding of HIV virus to human or animal cells
XX or by disturbing the 3D organisation of the virus.
XX
XX SQ Sequence 35 AA;
XX
XX Query Match 24.0%; Score 46; DB 16; Length 35;
XX Best Local Similarity 34.6%; Pred. No. 32;
XX Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;
XX
QY 8 TLVPGPGEPEFKT-----LRVONIG 27
Db 5 SIPIGPGRAFYTTGYKVKVKTIEPLG 30
XX
XX RESULT 7
XX AAR68712 ID AAR68712 standard; peptide; 33 AA.
XX
XX AC AAR68712;
XX
XX DT 07-SEP-1995 (first entry)
XX
XX DE T5/PRI chimera, CLTB-158.
XX
XX KW T-cell; epitope; HIV-1; core protein; p24E; B-cell; antigen;
XX gp160; gag; pol; vaccine; multimeric peptide; AIDS; 3D organisation.
XX
XX OS Human immunodeficiency virus type 1.
XX
XX FH Key Location/Qualifiers
XX FT 1..15 /label= T5
XX FT /note= "T cell epitope"
XX FT 17..33 /label= PRI
XX FT Peptide
```

```
FT /note= "B cell epitope"
XX
XX PN WO9429339-A.
XX
XX PD 22-DEC-1994.
XX
XX PF 08-JUN-1994; 94WO-CA00317.
XX
XX PR 09-JUN-1993; 93US-0073378.
XX
XX PA (CONN-) CONNAUGHT LAB LTD.
XX
XX PI Chong P, Klein MH, Sia CDY;
XX
XX DR WPI; 1995-036400/05.
XX
XX PT Novel tandem synthetic HIV-1 peptide(s) - comprising T-cell
XX epitope of gag protein linked to B-cell epitope of V3 loop
XX protein of an HIV-1 isolate
XX
XX PS Claim 4; Page 46; 69pp; English.
XX
XX CC This sequence represents the T-cell epitope derived from the HIV-1
XX core protein, T5, linked to consensus B-cell epitope, PRI which
XX represents a consensus sequence derived from New York and Amsterdam
XX HIV-1 isolates. Chimeric peptides such as this, may be used in the
XX production of HIV-1 vaccines. These peptide sequences may also be
XX used in the production of multimeric peptides in which the peptides
XX are C-terminally modified by the addition of a Lys residue which is
XX modified on its epsilon amino acid to carry an additional copy of
XX the peptide molecule. The linear and multimeric peptides may be
XX used for the treatment of AIDS by acting to displace the binding of
XX HIV virus to human or animal cells or by disturbing the 3D
XX organisation of the virus.
XX
XX SQ Sequence 33 AA;
XX
XX Query Match 23.7%; Score 45.5; DB 16; Length 33;
XX Best Local Similarity 34.6%; Pred. No. 36;
XX Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;
XX
QY 2 EVHPYG-----TLVPGPGEPEFKT 20
Db 7 KIEPLGVAPNTRKSIPIGPGRAFYTT 32
XX
XX RESULT 8
XX AAW25882 ID AAW25882 standard; peptide; 33 AA.
XX
XX AC AAW25882;
XX
XX DT 22-OCT-1997 (first entry)
XX
XX DE Chimeric T/B cell epitope peptide CLTB-158.
XX
XX KW HIV; human immunodeficiency virus; gag; T-cell; B-cell; epitope; env;
XX V3 loop; vaccine; determinant; chimeric.
XX
XX OS Synthetic.
XX
XX PN US5639854-A.
XX
XX PD 17-JUN-1997.
XX
XX PF 09-JUN-1993; 93US-0073378.
XX
XX PR 09-JUN-1994; 94US-0257528.
XX
XX PR 09-JUN-1993; 93US-0073378.
XX
XX PA (CONN-) CONNAUGHT LAB LTD.
XX
XX PI Chong P, Klein MH, Sia CDY;
```

```

XX DR WPI: 1997-332082/30.
XX
XX Tandem synthetic HIV peptide(s) useful as immunogens - comprising
PT gag protein T-cell epitope linked to env protein B-cell epitope
XX
XX Example 1; Column 23-24; 41pp; English.
XX
XX The invention relates to new synthetic peptides comprising at least one
CC amino acid sequence comprising an HIV gag protein T-cell epitope linked
CC at its C- or N-terminus to an amino acid sequence comprising a B-cell
CC epitope of the V3 loop of an HIV env protein, which can be used to
CC generate vaccines against HIV-1. The T-cell epitope sequence is pref.
CC selected from the T-helper determinant core peptides P24E, P24N, P24L,
CC P24W and P24H while the B-cell epitopes are derived from HIV strains
CC including CTLB-56, V3MN, CTLB-29, CTLB-55, SF2, LAI, IIB, RF, 26, 2054,
CC 1714 and BX08. The peptides are chimeric and can be linked to a
CC branched lys backbone. This sequence represents a chimaeric peptide
CC comprising the T-helper determinant core peptide T5 (AAW25881) with the
CC HIV-1 strain MN env protein V3 loop B-cell epitope PRI which corresponds
CC to a peptide generated from the consensus sequence of the New York and
CC Amsterdam HIV-1 isolates. The B-cell epitope is linked to the
CC C-terminus of the T-helper peptide.
XX
SQ Sequence 33 AA:
Query Match 23.7%; Score 45.5; DB 18; Length 33;
Best Local Similarity 34.6%; Pred. No. 36;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

OY 2 EVHPYG-----TLPGGPGPEPKTT 20
Db 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 9
AAW67398
ID AAW67398 standard; peptide; 33 AA.
XX
XX AAW67398;
XX
XX 25-JUN-1999 (first entry)
XX
XX HIV-1 peptide epitope CTLB-158.
XX
XX Immunogen; vaccine; HIV-1; T-cell; B-cell; epitope; core protein; gp120;
KW V3 loop.
XX
XX Synthetic.
OS Human immunodeficiency virus type 1.
XX
XX US5817754-A.
XX
XX 06-OCT-1998.
XX
XX 05-JUN-1995; 95US-0464329.
XX
XX 09-JUN-1994; 94US-0257528.
XX
XX 09-JUN-1993; 93US-0073378.
XX
XX 05-JUN-1995; 95US-0464329.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Chong P, Klein MH, Sia CDY;
XX
XX WPI: 1998-556461/47.
XX
XX Synthetic human immunodeficiency virus-1 peptide(s) - containing
PT T-cell epitope and B-cell epitope(s) are candidate vaccines against
XX HIV-1
XX
XX Disclosure; Column 23-24; 40pp; English.
XX

```

```

CC The invention relates to a novel immunogenic composition for use in
CC vaccines for the treatment of HIV-1 comprising an HIV-1-derived T-cell
CC epitope linked to an HIV-1-derived B-cell epitope. The T-cell epitopes
CC are generally designed based on the p24 core protein and the B-cell
CC epitopes from the V3 loop of the gp120 protein from various HIV-1
CC strains. This peptide corresponds to a fusion of the T5 epitope
CC (AAW67397) and the V3 loop B-cell epitope (AAW67365) which is a
CC consensus V3 loop peptide of the New York and Amsterdam HIV-1 isolates.
XX
SQ Sequence 33 AA:
Query Match 23.7%; Score 45.5; DB 19; Length 33;
Best Local Similarity 34.6%; Pred. No. 36;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

OY 2 EVHPYG-----TLPGGPGPEPKTT 20
Db 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 10
AAV39727
ID AAV39727 standard; peptide; 33 AA.
XX
XX AAV39727;
XX
XX 26-NOV-1999 (first entry)
XX
XX HIV1 chimeric peptide CTLB-158.
XX
XX HIV vaccine; immunogenic composition; T cell epitope; B cell epitope;
KW infection; antibody; antiviral.
XX
XX Human immunodeficiency virus type 1.
XX
XX US5951986-A.
XX
XX 14-SEP-1999.
XX
XX 06-JUN-1995; 95US-0467881.
XX
XX 09-JUN-1994; 94US-0257528.
XX
XX 09-JUN-1993; 93US-0073378.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Klein MH, Chong P, Sia CDY;
XX
XX WPI: 1999-550482/46.
XX
XX Immunogenic composition containing synthetic fusion polypeptides
PT containing both the T and B cell epitopes of the human immunodeficiency
PT virus, useful antigens in producing vaccines.
XX
XX Example 1; Column 25-26; 43pp; English.
XX
XX This sequence represents a fragment of a HIV1 protein, and can be used in
CC the immunogenic composition of the invention. The composition comprises a
CC synthetic fusion polypeptide which includes a sequence encoding 1 or more
CC T cell epitopes and a sequence encoding 1 or more B cell epitopes and a
CC carrier. Both the T cell and B cell epitopes are derived from HIV
CC proteins. The compositions are useful as vaccines against HIV infection.
CC The composition induces HIV-1-specific polyclonal antibodies that are
CC opsonising and antiviral. The peptide components may be selected to
CC induce a response against different viral isolates and in subjects who
CC recognise different T cell epitopes.
XX
SQ Sequence 33 AA:
Query Match 23.7%; Score 45.5; DB 20; Length 33;
Best Local Similarity 34.6%; Pred. No. 36;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

```





DE		Humant immunodeficiency virus peptide RPI50.
XX	KM	HIV isolate AVR-2; peptide RPI50; principal neutralising domain;
XX	KW	antibodies; diagnosis; prophylaxis; therapy; AIDS.
XX	XX	Synthetic.
XX	OS	
PN	PN	MO9003984-A.
XX	PD	19-APR-1990.
XX	PF	29-SEP-1989; 89WO-US04302.
XX	PR	19-SEP-1989; 89US-0407663, US-252949; WO-U04302.
XX	PA	(REPL-) REPLIGEN CORP.
PI	PI	Rusche JR, Putney SD, Javaherian K, Farley J, Grimalia R, Lynn D; Petro-Breyer J;
DR	XX	WPI; 1990-147824/19.
PT	PT	Principal neutralising domain of HIV variants - used for producing peptide(s) and antibodies for diagnosis; prophylaxis; and/or therapy of HIV infection.
PS	XX	Claim 8 (81); Page 77; 108pp; English.
CC	CC	Peptide RPI50 comprises segments of the Principal Neutralising Domain (envelope protein) from isolate AVR-2. The last Cys residue is added for the purpose of crosslinking to carrier proteins. Cysteine residues may be added, so that the residues at or near both ends form a disulfide bond, giving peptide a loop-like configuration, which can be utilised to enhance the immunogenic properties of the peptides. Protein is capable of eliciting, and/or binding with, neutralising Abs. The neutralising domain is bounded by cysteine residues which occur at positions 296 and 331. Peptides can be used as immunogens or screening reagents to generate or identify poly- or monoclonal antibodies. See also AAR04427-R04506 and AAQ04273-Q04279.
SO	Sequence	24 AA:
OY	Query Match	22.9%; Score 44; DB 11; Length 24;
	Best Local Similarity	50.0%; Pred. No. 41;
D8	Matches 9; Conservative	3; Mismatches 4; Indels 2; Gaps 1.
	11 VGGPEFEKTLRVQNLCG 28	
	:	
	9 IGPCRAFTTGR-IIGC 24	
RESULT 14		
AAB10706	ID	AAB10706 standard; Protein; 35 AA.
AC	AA	AAB10706;
DT	26-JAN-2001	(first entry)
DB	HIV-1 isolate PI-932 gp120 protein V3 loop region fragment.	
XX	Viral protein; vaccine; anti-viral; anti-HIV; therapy; infection;	
KW	gp120; V3 loop.	
XX	Human immunodeficiency virus type 1.	
OS		
PN	WO200047223-A2.	
PD	17-AUG-2000.	
PF	03-DEC-1999; 99WO-EP09759.	
PR	12-FEB-1999; 99DE-1007485.	

XX	(STRA-) STRATHMANN & CO AG.
PA	Schreiber M;
XX	WI, 2000-549084/50.
DR	Viral vaccine comprises a mixture of protein sequence variants of a
PT	single viral protein, which is useful for prevention and therapy of
PT	viral infections, especially HIV, in humans
XX	Disclosure; Fig 1; 79pp; German.
PS	This invention describes a novel protein vaccine which comprises a
XX	mixture of viral proteins, characterized in that the molecules are
CC	sequence variants of a single viral (partial) protein. The products of
CC	the invention have anti-viral and anti-HIV activity. Mixtures of
CC	structurally different viral proteins, that are sequence variants of a
CC	single protein are useful for production of vaccines for the prevention
CC	and/or therapy of viral infections in humans. The vaccines are especially
CC	useful for prevention and/or therapy of human immunodeficiency virus
CC	(HIV) infection in humans. The vectors and host cells are useful for
CC	expression of the protein/DNA mixtures, which are also useful for
CC	prevention and/or therapy of viral infection. This sequence represents
CC	a gp120 protein V3 loop fragment from HIV-1 isolate PI-932.
XX	Sequence 35 AA;
SQ	Query Match 22.9%; Score 44; DB 21; Length 35;
	Best Local Similarity 30.8%; Pred.No. 61;
	Matches 8; Conservative 7; Mismatches 7; Indels 4; Gaps 1;
OY	3 VHPYGT---LPVGPGEPEFKTLRVQ 24
	: ::     :::
Db	2 IRPHNTVDRIHIGGRSFTTRKIK 27
RESULT 15	
ID	AAR68710 standard; peptide; 33 AA.
AC	AAR68710;
DT	07-SEP-1995 (first entry)
DE	PRI/P24H chimera, CLTB-157.
XX	T-cell; epitope; HIV-1, core protein, p24E; B-cell; antigen;
KW	gp160; gag; pol; vaccine; multimeric peptide; AIDS; 3d organisation.
XX	Human immunodeficiency virus type 1.
OS	
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..18
FT	/label= PRI
FT	/note= "B cell epitope"
FT	19..33
FT	/label= P24H
FT	/note= "T cell epitope"
XX	
FN	MO9429339-A.
PD	22-DEC-1994.
XX	
PF	08-JUN-1994; 94WO-CA00317.
PR	09-JUN-1993; 93US-0073378.
PA	(CONN-) CONNAUGHT LAB LTD.
XX	
PI	Chong P, Klein MH, Sia CDY;
RR	WI, 1995-036400/05.

XX Novel tandem synthetic HIV-1 peptide(s) - comprising T-cell  
PT epitope of gag protein linked to B-cell epitope of V3 loop  
PT protein of an HIV-1 isolate  
XX  
XX

PS Claim 4; Page 45; 69pp; English.  
XX

CC This sequence represents the T-cell epitope derived from the HIV-1  
CC core protein, P24H, linked to a B-cell epitope, PRI which represents  
CC a consensus sequence derived from New York and Amsterdam HIV-1 isolates.  
CC Chimeric peptides such as this, may be used in the production of HIV-1  
CC vaccines. These peptide sequences may also be used in the production of  
CC multimeric peptides in which the peptides are C-terminally modified by  
CC the addition of a lys residue which is modified on its epsilon amino  
CC acid to carry an additional copy of the peptide molecule. The linear  
CC and multimeric peptides may be used for the treatment of AIDS by acting  
CC to displace the binding of HIV virus to human or animal cells or by  
CC disturbing the 3D organisation of the virus.  
XX

SQ Sequence 33 AA;

Query Match 22.7%; Score 43.5; DB 16; Length 33;

Best Local Similarity 50.0%; Pred. No. 68;  
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

OY 8 TLPGGPGPEFKTT-LRVONTL 26  
:::|||||

DB 5 SIPIGPGRAFTTGPVONI 24  
|||:

Search completed: July 16, 2003, 08:22:45  
Job time : 69 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:24:44 ; Search time 51 Seconds  
(without alignments)  
81.502 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192  
Sequence: 1 YEHVGYTLVPGPGPEFTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 113972

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pdp.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pdp.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pdp.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pdp.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pdp.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pdp.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pdp.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pdp.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pdp.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pdp.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pdp.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pdp.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pdp.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	21.9	29	9	US-10-106-698-7462
2	39	20.3	20	10	US-09-864-761-39770
3	39	20.3	21	10	US-09-864-761-46555
4	39	20.3	23	9	US-09-843-676-208
5	39	20.3	23	9	US-09-438-486-208
6	39	20.3	23	9	US-10-053-758-208
7	39	20.3	23	9	US-10-054-295-208
8	39	20.3	23	9	US-10-054-611-208
9	39	20.3	24	9	US-09-798-889-167
10	38	19.8	15	10	US-09-810-310-27
11	38	19.8	24	9	US-10-023-282-472
12	38	19.8	29	9	US-10-045-465-12
13	38	19.8	33	9	US-10-272-339A-18
14	38	19.8	33	9	US-10-272-339A-20
15	36.5	19.0	28	9	US-09-892-877-293
16	36.5	19.0	28	9	US-09-948-783-305
17	36	18.8	15	10	US-09-810-310-28
18	36	18.8	17	9	US-09-964-201A-3
19	36	18.8	22	9	US-09-932-613-167

20	36	18.8	25	9	US-10-050-882-98	Sequence 98, Appl
21	36	18.8	30	9	US-10-106-698-7473	Sequence 7473, Ap
22	36	18.8	33	9	US-10-106-698-7486	Sequence 7486, Ap
23	35.5	18.5	21	9	US-10-097-065-606	Sequence 606, App
24	35.5	18.5	34	9	US-10-106-698-7549	Sequence 7549, Ap
25	35	18.2	12	9	US-10-254-446A-193	Sequence 193, App
26	35	18.2	15	10	US-09-767-460-7	Sequence 7, Appl
27	35	18.2	15	10	US-09-810-310-30	Sequence 30, Appl
28	35	18.2	25	10	US-09-803-165-17	Sequence 17, Appl
29	35	18.2	28	10	US-09-864-761-41533	Sequence 41533, A
30	35	18.2	29	10	US-09-864-761-35300	Sequence 35300, A
31	35	18.2	31	10	US-09-864-761-43565	Sequence 43565, A
32	35	18.2	33	10	US-09-864-761-44246	Sequence 44246, A
33	34.5	18.0	21	9	US-10-106-698-8032	Sequence 8032, Ap
34	34.5	18.0	23	9	US-10-106-698-8221	Sequence 8221, Ap
35	34	17.7	10	9	US-09-087-513-18	Sequence 18, Appl
36	34	17.7	10	9	US-09-810-310-20	Sequence 20, Appl
37	34	17.7	17	9	US-10-059-271-18	Sequence 18, Appl
38	34	17.7	18	10	US-09-030-619-126	Sequence 126, App
39	34	17.7	19	9	US-10-059-271-33	Sequence 33, Appl
40	34	17.7	19	10	US-09-308-511-1	Sequence 1, Appl
41	34	17.7	20	9	US-10-062-831-203	Sequence 203, App
42	34	17.7	21	10	US-09-853-830-49	Sequence 49, Appl
43	34	17.7	28	9	US-09-462-713-7	Sequence 7, Appl
44	34	17.7	30	9	US-10-040-862-10444	Sequence 10444, A
45	34	17.7	30	9	US-10-023-282-1093	Sequence 1093, Ap

#### ALIGNMENTS

RESULT 1  
US-10-106-698-7462  
; Sequence 7462, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: PA00501  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 7462  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (3)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids.  
; NAME/KEY: MISC FEATURE  
; LOCATION: (6)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC FEATURE  
; LOCATION: (7)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC FEATURE  
; LOCATION: (9)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; US-10-106-698-7462

Query Match 21.9%; Score 42; DB 9; Length 29;  
Best Local Similarity 56.2%; Pred. No. 45;  
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;  
QY 4 HPGYTL--PVGPGRPF 17

Db 11 HYFGKLVPPAGTGPFF 26

## RESULT 2

US-09-864-761-39770  
Sequence 39770, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 39770  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005747.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.72  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.79  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
OTHER INFORMATION: EST\_HUMAN HIT: H87835.1, EVALUATE 2.00e-06  
US-09-864-761-39770

Query Match 20.3%; Score 39; DB 10; Length 20;

Best Local Similarity 58.3%; Pred. No. 77;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

## Query 4 HPYGTLPVGPFP 15

Db 1 HSWGTAKVGPFP 12

## RESULT 3

US-09-864-761-46555  
Sequence 46555, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 46555  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC009973.2  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
OTHER INFORMATION: EST\_HUMAN HIT: AL121516.1, EVALUATE 8.00e-07

US-09-864-761-46555

Query Match	20.3%	Score 39;	DB 10;	Length 21;
Best Local Similarity	57.9%;	Pred. No. 81;		
Matches	11;	Conservative	1;	Mismatches 3;
				Indels 4;
				Gaps 2

Db 6 DVH-RGTL--PGPNHKTT 20

RESULT 4  
US-09-843-676-208

APPLICANT: Cech, Thomas R.

TITLE OF INVENTION: No. US20020164786A1el Telomerase  
NUMBER OF SEQUENCES: 225

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/843,676  
FILING DATE: 26-Apr-2001  
CLASSIFICATION: 536

APPLICATION NUMBER: US/08/854,055  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/844,413  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300

SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid

SEQUENCE	DESCRIPTION:	SEQ ID NO:	208:
US-09-843-676-208			

Query Match	20.3%	Score 39;	DB 9;	Length 23;
Best Local Similarity	58.3%;	Pred. No. 90;		
Matches 7;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0

Db 9 HPRENL PQDPGR 20

RESULT 5  
US-09-438-486-208

APPLICANT: Cech, Thomas R.  
 APPLICANT: Lingner, Joachim  
 APPLICANT: Nakamura, Toru  
 APPLICANT: Chapman, Karen B.  
 APPLICANT: Martin, Gregg, B.  
 APPLICANT: Hatley, Calvin  
 APPLICANT: Andrews, William H.  
 TITLE OF INVENTION: No. US20030009019A1 Telomerase  
 NUMBER OF SEQUENCES: 223

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent'n Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/438,486  
FILING DATE: 12-NOV-1999

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997

CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002931US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 20

LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS:

Query Match

4	HPYGLTVGPGP	15	Indels	0	Gaps	0
7	Conservative	0	Mismatches	5		
30	Matches	30				
30	Indels	30				
30	Gaps	30				
30	Mismatches	30				

RESULT 6  
US-10-053-758-208  
Sequence 208, Application US/10053758  
Publication No. US20030032075A1  
GENERAL INFORMATION:  
APPLICANT: Cecch, Thomas R.  
Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Morin, Gregg B.  
Harley, Calvin  
Andrews, William H.  
TITLE OF INVENTION: No. US20030032075A1el Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/053,758  
FILING DATE: 18-Jan-2002  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/854,050  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002930US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0300  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 208:  
US-10-053-758-208  
Query Match 20.3%; Score 39; DB 9; Length 23;  
Best Local Similarity 58.3%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 4 HPGTLPVGP 15  
DB 9 HPRNLPDPP 20  
RESULT 7  
US-10-054-295-208  
Sequence 208, Application US/10054295  
Publication No. US20030044953A1  
GENERAL INFORMATION:

APPLICANT: Cecch, Thomas R.  
Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Morin, Gregg B.  
Harley, Calvin  
Andrews, William H.  
TITLE OF INVENTION: No. US20030044953A1el Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/054,295  
FILING DATE: 18-Jan-2002  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/854,050  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002930US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0300  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 208:  
US-10-054-295-208  
Query Match 20.3%; Score 39; DB 9; Length 23;  
Best Local Similarity 58.3%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 4 HPGTLPVGP 15  
DB 9 HPRNLPDPP 20  
RESULT 8  
US-10-054-611-208  
Sequence 208, Application US/10054611  
Publication No. US20030059787A1  
GENERAL INFORMATION:  
APPLICANT: Cecch, Thomas R.  
Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Morin, Gregg B.  
Harley, Calvin  
Andrews, William H.  
TITLE OF INVENTION: No. US20030059787A1el Telomerase

NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/054,611  
FILING DATE: 18-Jan-2002  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/854,050  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002930US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 208:  
US-10-054-611-208  
Query Match 20.3%; Score 39; DB 9; Length 23;  
Best Local Similarity 58.3%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 4 HRYGLPVGPGP 15  
DB 9 HRENLPQDPGP 20  
RESULT 9  
US-09-798-889-167  
Sequence 167, Application US/09798889  
Publication No. US20030004324A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 31 Human secreted proteins  
FILE REFERENCE: P2026P1  
CURRENT APPLICATION NUMBER: US/09/798,889  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/393,022  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-09  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,714  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,686  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,687  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,656  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12  
NUMBER OF SEQ ID NOS: 185

SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 167  
LENGTH: 24  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-798-889-167  
Query Match 20.3%; Score 39; DB 9; Length 24;  
Best Local Similarity 54.5%; Pred. No. 94;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 YEHPYGLPV 11  
DB 7 YTKHPFGHPV 17  
RESULT 10  
US-09-810-310-27  
Sequence 27, Application US/09810310  
Patent No. US20020044948A1  
GENERAL INFORMATION:  
APPLICANT: Khleif, Samir N.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CO-STIMULATION OF  
FILE REFERENCE: 15280-41510US  
CURRENT APPLICATION NUMBER: US/09/810,310  
PRIOR FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 60/189,396  
PRIOR FILING DATE: 2000-03-15  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 27  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: HIV PEPTIDE  
US-09-810-310-27  
Query Match 19.8%; Score 38; DB 10; Length 15;  
Best Local Similarity 46.2%; Pred. No. 76;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 8 TLPVGPGEFXT 20  
DB 1 SIHIGPAPFYAT 13  
RESULT 11  
US-10-023-282-472  
Sequence 472, Application US/10023282  
Publication No. US20030092893A1  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: P2007P1  
CURRENT APPLICATION NUMBER: US/10/023,282  
PRIOR FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 09/205,258  
PRIOR FILING DATE: 1998-12-04  
PRIOR APPLICATION NUMBER: PCT/US98/11422  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/048,885  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/049,375  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,881  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,880  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/048,896

```

; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 472
; LENGTH: 24

```

---

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (24)
; OTHER INFORMATION: Xaa equals stop translation
US-10-023-282-472

Query Match
Best Local Similarity 19.8%; Score 38; DB 9; Length 24;
Matches 8; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 4 HPVGLPVGP 15
DB 13 HPPG--PGPGP 22

RESULT 12
US-10-045-465-12
; Sequence 12, Application US/10045465
; Publication No. US20030049605A1
; GENERAL INFORMATION:
; APPLICANT: Van Es, Helmut
; TITLE OF INVENTION: Display of viral proteins
; FILE REFERENCE: 2183-4079US
; CURRENT APPLICATION NUMBER: US/10/045,465
; CURRENT FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: 09/315,244
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: EP 98201678.4
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1) (29)
; OTHER INFORMATION: /No. US20030049605A1="Amino acid sequence of ecotropic env varia
; OTHER INFORMATION: and/or parts;
; OTHER INFORMATION: COOH-terminus of construct 3 at location 253-266
US-10-045-465-12

Query Match
Best Local Similarity 19.8%; Score 38; DB 9; Length 29;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 21 LRVNLTGCVVSG 33
DB 4 LRVNLTGPRVTS 16

RESULT 13
US-10-272-339A-18
; Sequence 18, Application US/10272339A
; Publication No. US20030120061A1
; GENERAL INFORMATION:
; APPLICANT: Jingwu, Zhang Z.
; TITLE OF INVENTION: T Cell Receptor VB-DB-JB Sequence and Methods For Its
; FILE REFERENCE: 21838-00029
; CURRENT APPLICATION NUMBER: US/10/272,339A
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 09/641,576
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 09/507,819
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/121,311
; PRIOR FILING DATE: 1999-02-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.1

```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:21:34 ; Search time 26 Seconds  
(without alignments)  
39,608 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192  
Sequence: 1 YEHPYGTLPVGPPEFKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 167989

Minimum DB seq length: 0  
Maximum DB seq length: 35

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	24.0	33	1	US-08-257-528B-84 Sequence 84, Appl
2	46	24.0	33	1	US-08-460-602A-84 Sequence 84, Appl
3	46	24.0	33	1	US-08-463-966A-84 Sequence 84, Appl
4	46	24.0	33	1	US-08-465-217A-84 Sequence 84, Appl
5	46	24.0	33	2	US-08-464-329A-84 Sequence 84, Appl
6	46	24.0	33	2	US-08-462-507A-84 Sequence 84, Appl
7	46	24.0	33	2	US-08-467-881A-84 Sequence 84, Appl
8	46	24.0	34	4	US-09-141-833-11 Sequence 11, Appl
9	45.5	23.7	33	1	US-08-257-528B-83 Sequence 83, Appl
10	45.5	23.7	33	1	US-08-460-602A-83 Sequence 83, Appl
11	45.5	23.7	33	1	US-08-463-966A-83 Sequence 83, Appl
12	45.5	23.7	33	1	US-08-465-217A-83 Sequence 83, Appl
13	45.5	23.7	33	2	US-08-464-329A-83 Sequence 83, Appl
14	45.5	23.7	33	2	US-08-462-507A-83 Sequence 83, Appl
15	45.5	23.7	33	2	US-08-467-881A-83 Sequence 83, Appl
16	45	23.4	35	3	US-08-513-968-13 Sequence 13, Appl
17	43.5	22.7	33	1	US-08-257-528B-81 Sequence 81, Appl
18	43.5	22.7	33	1	US-08-460-602A-81 Sequence 81, Appl
19	43.5	22.7	33	1	US-08-463-966A-81 Sequence 81, Appl
20	43.5	22.7	33	1	US-08-465-217A-81 Sequence 81, Appl
21	43.5	22.7	33	2	US-08-464-329A-81 Sequence 81, Appl
22	43.5	22.7	33	2	US-08-462-507A-81 Sequence 81, Appl
23	43.5	22.7	33	2	US-08-467-881A-81 Sequence 81, Appl
24	43	22.4	18	1	US-08-257-528B-50 Sequence 50, Appl
25	43	22.4	18	1	US-08-460-602A-50 Sequence 50, Appl
26	43	22.4	18	1	US-08-463-966A-50 Sequence 50, Appl
27	43	22.4	18	1	US-08-465-217A-50 Sequence 50, Appl

28	43	22.4	18	2	US-08-464-329A-50 Sequence 50, Appl
29	43	22.4	18	2	US-08-462-507A-50 Sequence 50, Appl
30	43	22.4	18	2	US-08-467-881A-50 Sequence 50, Appl
31	43	22.4	23	3	US-08-513-968-71 Sequence 71, Appl
32	43	22.4	23	3	US-08-513-968-72 Sequence 72, Appl
33	43	22.4	33	1	US-08-257-528B-41 Sequence 41, Appl
34	43	22.4	33	1	US-08-257-528B-80 Sequence 41, Appl
35	43	22.4	33	1	US-08-460-602A-41 Sequence 41, Appl
36	43	22.4	33	1	US-08-460-602A-80 Sequence 41, Appl
37	43	22.4	33	1	US-08-463-966A-41 Sequence 41, Appl
38	43	22.4	33	1	US-08-463-966A-80 Sequence 41, Appl
39	43	22.4	33	1	US-08-465-217A-41 Sequence 41, Appl
40	43	22.4	33	1	US-08-465-217A-80 Sequence 41, Appl
41	43	22.4	33	2	US-08-464-329A-41 Sequence 41, Appl
42	43	22.4	33	2	US-08-464-329A-80 Sequence 41, Appl
43	43	22.4	33	2	US-08-462-507A-41 Sequence 41, Appl
44	43	22.4	33	2	US-08-462-507A-80 Sequence 41, Appl
45	43	22.4	33	2	US-08-467-881A-41 Sequence 41, Appl

#### ALIGNMENTS

RESULT 1  
US-08-257-528B-84  
Sequence 84, Application US/08257528B  
Patent No. 5639854  
GENERAL INFORMATION:  
APPLICANT: SIA, Charles D.Y.  
APPLICANT: CHONG, Pele  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides  
NUMBER OF SEQUENCES: 101  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/257,528B  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-336 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1163  
TELEFAX: (416) 595-1155  
INFORMATION FOR SEQ ID NO: 84:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-257-528B-84

Query Match 24.0%; Score 46; DB 1; Length 33;  
Best Local Similarity 34.6%; Pred. No. 4.1;  
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

CY 8 TLVPGPGPEFKT-----LRVQNLG 27  
DB 5 SIPIGGRAPFTTGYKRVVKIEPLG 30

```
RESULT 2
US-08-460-602A-84
; Sequence 84, Application US/08460602A
; Patent No. 5759769
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,602A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,528
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,378
; FILING DATE: 09-JUN-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-450 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-460-602A-84
;
Query Match      24.0%; Score 46; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.1;
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY      8 TLVPGRGPFFKT-----LRVQNLG 27
Db      5 SIPIGGRGYTTGYKRVVKIEPLG 30

RESULT 3
US-08-463-966A-84
; Sequence 84, Application US/08463966A
; Patent No. 5795955
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
```

```

; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,966A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,528
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,378
; FILING DATE: 09-JUN-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-487 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-463-966A-84
;
Query Match      24.0%; Score 46; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.1;
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY      8 TLVPGRGPFFKT-----LRVQNLG 27
Db      5 SIPIGGRGYTTGYKRVVKIEPLG 30

RESULT 4
US-08-465-217A-84
; Sequence 84, Application US/08465217A
; Patent No. 5800822
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,217A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,528
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
```



```
; Sequence 84, Application US/08467881A
; Patent No. 5951986
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,881A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,528
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,378
; FILING DATE: 09-JUN-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-488 MIS:jfb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-467-881A-84

Query Match      24.0%; Score 46; DB 2; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.1;
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY      8 TLVPGGPEFKTT-----LRVONG 27
DB      5 SLIPGGRAPFTTGYKVKVKKLPLG 30

RESULT 8
US-09-141-833-11
; Sequence 11, Application US/09141833
; Patent No. 6168784
; GENERAL INFORMATION:
; APPLICANT: OFFORD, ROBIN E
; APPLICANT: THOMPSON, DARREN
; APPLICANT: WILKEN, JILL
; TITLE OF INVENTION: N-TERMINAL MODIFICATIONS OF RANTES AND METHODS OF USE
; FILE REFERENCE: GREN-026/03US
; CURRENT APPLICATION NUMBER: US/09/141,833
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: 60/056,292
; EARLIER FILING DATE: 1997-09-03
; EARLIER APPLICATION NUMBER: 60/077,874
; EARLIER FILING DATE: 1998-03-13
; EARLIER APPLICATION NUMBER: 60/090,834
```

```
; EARLIER FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO. 11
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-141-833-11

Query Match      24.0%; Score 46; DB 4; Length 34;
Best Local Similarity 47.4%; Pred. No. 4.3;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY      8 TLVPGGPEFKTTLRVONL 26
DB      11 SLIPGGRAPFTTGYKVKVKKLPLG 29

RESULT 9
US-08-257-528B-83
; Sequence 83, Application US/08257528B
; Patent No. 5639854
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,528B
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-336 MIS:jfb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-528B-83

Query Match      23.7%; Score 45.5; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.9;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY      2 EVHPYG-----TLVPGGPEFKTT 20
DB      7 KIEPLGVAENPKRSIPIGGRAPFTT 32

RESULT 10
US-08-460-602A-83
; Sequence 83, Application US/08460602A
; Patent No. 5759769
; GENERAL INFORMATION:
```

APPLICANT: SIA, Charles D.Y.  
APPLICANT: CHONG, Pele  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides  
NUMBER OF SEQUENCES: 101  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,602A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,528  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/073,378  
FILING DATE: 09-JUN-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-450 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-460-602A-83

Query Match 23.7%; Score 45.5; DB 1; Length 33;  
Best Local Similarity 34.6%; Pred. No. 4.9;  
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

Qy 2 EVHPYG-----TLFVGSGPEPKTT 20  
Db 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 11  
US-08-463-966A-83  
Sequence 83, Application US/08463966A  
Patent No. 5795955  
GENERAL INFORMATION:  
APPLICANT: SIA, Charles D.Y.  
APPLICANT: CHONG, Pele  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides  
NUMBER OF SEQUENCES: 101  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,966A  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,528  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/073,378  
FILING DATE: 09-JUN-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-487 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-463-966A-83

Query Match 23.7%; Score 45.5; DB 1; Length 33;  
Best Local Similarity 34.6%; Pred. No. 4.9;  
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

Qy 2 EVHPYG-----TLFVGSGPEPKTT 20  
Db 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 12  
US-08-465-217A-83  
Sequence 83, Application US/08465217A  
Patent No. 5800822  
GENERAL INFORMATION:  
APPLICANT: SIA, Charles D.Y.  
APPLICANT: CHONG, Pele  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides  
NUMBER OF SEQUENCES: 101  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,217A  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,528  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/073,378  
FILING DATE: 09-JUN-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-486 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-465-217A-83

Query Match 23.7%; Score 45.5; DB 1; Length 33;  
Best Local Similarity 34.6%; Pred. No. 4.9;  
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 2 EVHPYG-----TLVPGPGPEPKTT 20  
DB 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 13  
US-08-464-329A-83  
Sequence 83, Application US/08464329A  
Patent No. 5817754  
GENERAL INFORMATION:  
APPLICANT: SIA, Charles D.Y.  
APPLICANT: CHONG, Pele  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides  
NUMBER OF SEQUENCES: 101  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,329A  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,528  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/073,378  
FILING DATE: 09-JUN-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-449 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-464-329A-83

Query Match 23.7%; Score 45.5; DB 2; Length 33;

Best Local Similarity 34.6%; Pred. No. 4.9;  
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 2 EVHPYG-----TLVPGPGPEPKTT 20  
DB 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 14  
US-08-462-507A-83  
Sequence 83, Application US/08462507A  
Patent No. 5876731  
GENERAL INFORMATION:  
APPLICANT: SIA, Charles D.Y.  
APPLICANT: CHONG, Pele  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides  
NUMBER OF SEQUENCES: 101  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,507A  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,528  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/073,378  
FILING DATE: 09-JUN-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-451 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-462-507A-83

Query Match 23.7%; Score 45.5; DB 2; Length 33;  
Best Local Similarity 34.6%; Pred. No. 4.9;  
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 2 EVHPYG-----TLVPGPGPEPKTT 20  
DB 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 15  
US-08-467-881A-83  
Sequence 83, Application US/08467881A  
Patent No. 5951986  
GENERAL INFORMATION:  
APPLICANT: SIA, Charles D.Y.  
APPLICANT: CHONG, Pele



APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides  
NUMBER OF SEQUENCES: 101  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,881A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,528  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/073,378  
FILING DATE: 09-JUN-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, MICHAEL I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-488 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-881A-83

Query Match 23.7%; Score 45.5; DB 2; Length 33;  
Best Local Similarity 34.6%; Pred. No. 4.9;  
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 2 EVHPYG-----TLPVGGPGEFFKT 20  
Db 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

Search completed: July 16, 2003, 08:25:59  
Job time : 27 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:52 ; Search time 1.1294 Seconds

(without alignments)  
1872.646 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114

Sequence: 1 KLGFPAHKKIPEEEKREKLEQ 22

Scoring table: BLOSUM62

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	46.5	214	2 A11073	probable membrane
2	51	44.7	672	2 G69503	signal-transducing
3	51	44.7	856	2 H64552	endopeptidase Clp
4	51	44.7	1045	2 S60571	integrin alpha v c
5	50	43.9	357	2 AC1104	B. subtilis YacI p
6	50	43.9	505	1 S77034	protein kinase pkn
7	50	43.9	1034	2 A36108	integrin alpha-v c
8	50	43.9	1044	2 T10050	integrin alpha-v c
9	50	43.9	1048	2 A27421	integrin alpha-5 c
10	49.5	43.4	236	2 T12766	probable lipoprote
11	49.5	43.4	959	1 B60017	outer capsid prote
12	49	43.0	357	2 AC1466	B. subtilis YacI p
13	49	43.0	655	2 T06108	hypothetical prote
14	49	43.0	853	2 S74279	hypothetical prote
15	49	43.0	856	2 C71856	probable endopepti
16	48.5	42.5	407	2 S23325	M2 protein precurs
17	48	42.1	421	2 H90433	hypothetical prote
18	48	42.1	573	2 A55767	squalene monooxyge
19	48	42.1	962	2 JCS5808	G protein-coupled
20	48	42.1	1464	2 T13716	bazooka gene prote
21	48	42.1	1526	2 JN0598	DNA topoisomerase
22	47	41.2	102	2 F75077	hypothetical prote
23	47	41.2	112	2 E70433	flagellar switch p
24	47	41.2	206	2 T16946	hypothetical prote
25	47	41.2	512	2 G97612	probable transpos
26	47	41.2	512	2 G98236	probable transpos
27	47	41.2	512	2 AC3303	IS3 family transpo
28	47	41.2	512	2 AD2835	IS3 family transpo
29	47	41.2	512	2 AD3049	IS3 family transpo

30	47	41.2	692	2 S37976	hypothetical prote
31	47	41.2	861	2 T02267	trehalose-6-phosph
32	46	40.4	278	2 D84492	hypothetical prote
33	46	40.4	308	2 S11153	oligopeptide trans
34	46	40.4	308	2 E95220	hypothetical prote
35	46	40.4	323	2 E98084	hypothetical prote
36	46	40.4	450	2 S37900	hypothetical prote
37	46	40.4	544	2 AH2971	hypothetical prote
38	46	40.4	544	2 B96311	probable ATP-bind
39	46	40.4	847	2 C96703	hypothetical prote
40	46	40.4	1037	2 A60163	glycoprotein fib
41	46	40.4	1407	2 S28589	trichonhyalin - tab
42	45.5	39.9	570	2 F70332	proline-trna synth
43	45	39.5	273	2 T47612	hypothetical prote
44	45	39.5	325	2 T18283	hypothetical prote
45	45	39.5	333	2 T08850	alternative respir

## ALIGNMENTS

## RESULT 1

A11073

probable membrane protein smp [imported] - Salmonella enterica subsp. enterica serovar Typh

C/Species: Salmonella enterica subsp. enterica serovar Typh

A/Note: this species has also been called Salmonella typhi

C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #ext\_change 09-Nov-2001

C/Accession: A11073

R/Parhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, T., Connor, P., Cronin, A., Davies, P., Davies, R.M., Dowd, L., White, N., Farrar, S., Moul, S., O'Gaora, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K.

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero

A/Reference number: AB0502; PMID:11677608

A/Accession: A11073

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-214 <PAR>

A/Cross-references: GB:AL513382; PIDN:CMD03408.1; PID:g16505677; GSPDB:GN00176

A/Genetics:

A:gene: smp

Query Match

Best local similarity 71.4%; Pred. No. 5;

Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREK 19

DB 200 ASKPVEPEERREK 213

RESULT 2

G69503

signal-transducing histidine kinase homolog - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #ext\_change 21-Jul-2000

C/Accession: G69503

R/Klein, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson

., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirtress, E.F.

Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Urtreback, T., Cottrell, M.D., Spriggs, T., Artach, P., Kalne, B.P., Sykes, S.

Smith, H.O., Moese, C.R., Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A/Reference number: A69250; PMID:98049343; PMID:9389475

A/Accession: G69503

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-672 <KLE>

A/Cross-references: GB:AE000963; GB:AE000782; NID:g2689286; PIDN:AA089224.1; PID:g26485C

Query Match

44.7%; Score 51; DB 2; Length 672;

Best Local Similarity 55.6%; Pred. No. 30;  
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 5 FAHKKIPPEEKREKLEO 22  
DB 572 FAFKMEDEERREELKQ 589

## RESULT 3

H64552

endopeptidase Clp ATP-binding chain B - Helicobacter pylori (strain 26695)  
N/Alternate names: ATP-dependent Clp proteinase regulatory chain  
N/Contains: adenosinetriphosphatase (EC 3.6.1.3)  
C/Species: Helicobacter pylori

C/Date: 09-Aug-1997 #sequence\_revision 15-Aug-1997 #text\_change 19-Jan-2001  
C/Accession: H64552

R/TMD: J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Kalish, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Meldman, J.M.; Fujii, C.; Bowman, C.; Wathey, L.; Nature 388, 539-547, 1997

A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A/Reference number: A64520; PMID:97394467; PMID:9252185  
A/Accession: H64552

A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA

A/Residues: 1-856 <TOM>  
A/Cross-references: GB:AE000545; GB:AE000511; NID:92313349; PIDN:MAD07330.1; PID:g21335

C/Function: allows clp to hydrolyze polypeptides and proteins, probably by a chaperone activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller C/Superfamily: endopeptidase Clp ATP-binding chain

C/Keywords: ATP; duplication; hydrolase; molecular chaperone; nucleotide binding; P-loop  
F/204-211/Region: nucleotide-binding motif A (P-loop)  
F/211-276/Region: nucleotide-binding motif B  
F/605-612/Region: nucleotide-binding motif A (P-loop)  
F/673-678/Region: nucleotide-binding motif B  
F/210/Binding site: ATP (Lys) #status predicted  
F/611/Binding site: ATP (Lys) #status predicted

Query Match 44.7%; Score 51; DB 2; Length 856;  
Best Local Similarity 45.0%; Pred. No. 38;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 3 GFPAHKKIPPEEKREKLEO 22  
DB 496 GEIRYSKIPENKKEELOR 515

## RESULT 4

S60571

integrin alpha v chain precursor - Iberian ribbed newt  
C/Species: Pleurodeles waltl (Iberian ribbed newt)  
C/Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 29-Sep-1999

C/Accession: S60571  
R/Alfandari, D.; Whitaker, C.A.; Desimone, D.W.; Daribere, T.  
Dev. Biol. 170, 249-261, 1995

A/Title: Integrin alpha-(v) subunit is expressed on mesodermal cell surfaces during amphi  
A/Reference number: S60571; PMID:95377519; PMID:7649360  
A/Accession: S60571

A/Status: preliminary; nucleic acid sequence not shown  
A/Molecule type: mRNA

A/Residues: 1-1045 <ALF>  
A/Cross-references: EMBL:X81108; NID:g1008137; PIDN:CA57014.1; PID:g1008138

C/Superfamily: integrin alpha-2b chain  
F/1-29/Domain: signal sequence #status predicted <SIG>  
F/30-1045/Product: integrin alpha v chain #status predicted <MAT>

Query Match 44.7%; Score 51; DB 2; Length 1045;  
Best Local Similarity 42.9%; Pred. No. 47;  
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 KLGFAHKKIPPEEKREKLE 21

DB 1014 KFGFKRVPRPQEEGEREQLQ 1034

## RESULT 5

AC1104

B. subtilis yacI protein homolog lmo0234 [imported] - Listeria monocytogenes (strain EGI  
C/Species: Listeria monocytogenes  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C/Accession: AC1104  
R/Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001

A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; M-  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,

A/Title: Comparative genomics of Listeria species.  
A/Reference number: AB1077; PMID:21537279; PMID:11679669

A/Accession: AC1104  
A/Status: preliminary

A/Molecule type: DNA  
A/Residues: 1357 <GLA>

A/Cross-references: GB:NC\_003210; PIDN:CAD00761.1; PID:g16409599; GSPDB:GN00177  
A/Experimental source: strain EGD-e  
C/Genetics:

A/Gene: lmo0234  
C/Superfamily: conserved hypothetical protein yacI

Query Match 43.9%; Score 50; DB 2; Length 357;  
Best Local Similarity 71.4%; Pred. No. 22;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 6 AHKKIPPEEKREK 19  
DB 146 AKKRTPEEKREK 159

## RESULT 6

S77034

protein kinase pknA (EC 2.7.1.-), 55K - Synechocystis sp. (strain PCC 6803)  
N/Alternate names: protein sl10776  
C/Species: Synechocystis sp.

A/Variety: PCC 6803  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C/Accession: S77034  
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asami, E.; Nakamura, Y.; Miyajima, N.;  
o, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,

DNA Res. 3, 109-136, 1996  
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

s.  
A/Reference number: S74322; PMID:97061201; PMID:8905231  
A/Accession: S77034

A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA

A/Residues: 1-505 <KAN>  
A/Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BA10726.1; PID:g10065

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C/Genetics:

A/Gene: pknA  
C/Superfamily: Synechocystis sp. protein kinase pknA, 55K; protein kinase homology  
C/Keywords: phosphotransferase; protein kinase

F/7-268/Domain: protein kinase homology <KIN>

Query Match 43.9%; Score 50; DB 1; Length 505;  
Best Local Similarity 58.8%; Pred. No. 31;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 6 AHKKIPPEEKREKLEO 22  
DB 376 AEKLIANEKRORELEQ 392

## RESULT 7

A36108  
 Integrin alpha-V chain precursor - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 18-Jan-1991 #sequence\_revision 18-Jan-1991 #text\_change 29-Sep-1999  
 C:Accession: A36108  
 R:Bossy, B.; Reichardt, L.F.  
 Biochemistry 29, 10191-10198, 1990  
 A:Title: Chick integrin alpha-v subunit molecular analysis reveals high conservation of  
 A:Reference number: A36108; PMID:31104936; PMID:1703004  
 A:Accession: A36108  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1034 <BOS>  
 A:Cross-references: GB:M60517; NID:G212876; PIDN:AAA49138.1; PID:G212877  
 C:Superfamily: Integrin alpha-2b chain  
 C:Keywords: cell adhesion; cytoskeleton; transmembrane protein

Query Match 43.9%; Score 50; DB 2; Length 1034;  
 Best Local Similarity 38.1%; Pred. No. 64;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db 1 KLGFFAKKIPPEEKREKLE 21  
 1003 RMGFFKVRPPQEEQEREQLQ 1023

RESULT 8  
 T10050  
 Integrin alpha-v chain precursor - mouse  
 N/Alternate names: vitronectin receptor alpha chain  
 C:Species: Mus musculus (house mouse)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
 C:Accession: T10050  
 R:Wada, J.; Kumar, A.; Liu, Z.; Ruoslahti, E.; Reichardt, L.; Marvaldi, J.; Kanar, Y.S.  
 J. Cell Biol. 133, 1161-1176, 1996  
 A:Title: Cloning of mouse integrin alpha v cDNA and role of the alpha v-related matrix  
 A:Reference number: Z16920; PMID:96176309; PMID:8601592  
 A:Accession: T10050  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1044 <WAD>  
 A:Cross-references: EMBL:U01415; NID:G537490; PIDN:MAC52497.1; PID:G537491  
 A:Experimental source: strain CD-1, kidney  
 C:Superfamily: Integrin alpha-2b chain  
 C:Keywords: cell adhesion; receptor

Query Match 43.9%; Score 50; DB 2; Length 1044;  
 Best Local Similarity 38.1%; Pred. No. 64;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db 1 KLGFFAKKIPPEEKREKLE 21  
 1013 RMGFFKVRPPQEEQEREQLQ 1033

RESULT 9  
 A27421  
 Integrin alpha-5 chain precursor - human  
 N/Alternate names: CD51 antigen; serum spreading factor; vitronectin receptor alpha chain  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 31-Dec-2000  
 C:Accession: A27421; S47541; B29418; A26482; A32287; A30298; A35035  
 R:Suzuki, S.; Argaves, W.S.; Arat, H.; Languino, L.R.; Pierschbacher, M.D.; Ruoslahti, J.  
 J. Biol. Chem. 262, 14080-14085, 1987  
 A:Title: Amino acid sequence of the vitronectin receptor alpha subunit and comparative  
 A:Reference number: A27421; PMID:88007656; PMID:2443500  
 A:Accession: A27421  
 A:Molecule type: mRNA  
 A:Residues: 1-1048 <SUZ>  
 A:Cross-references: GB:M14648; GB:J02826; GB:M18365; NID:G340306; PIDN:AAA36808.1; PID:G  
 R;Donahue, J.P.; Sug9, N.; Hawiger, J.  
 Biochim. Biophys. Acta 1219, 228-232, 1994  
 A:Title: The integrin alpha(v) gene: identification and characterization of the promoted

A:Reference number: S47541; PMID:94368864; PMID:7522056  
 A:Accession: S47541  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-61 <DON>  
 A:Cross-references: EMBL:U07375  
 R:Fitzeveld, L.A.; Poncz, M.; Steiner, B.; Rall Jr., S.C.; Bennett, J.S.; Phillips, D.  
 Biochemistry 26, 8158-8165, 1987  
 A:Title: Comparison of CDNA-derived protein sequences of the human fibronectin and vitr  
 A:Reference number: A90526; PMID:88163472; PMID:2450560  
 A:Accession: B29418  
 A:Molecule type: mRNA  
 A:Residues: 1-433 <FIT>  
 R:Suzuki, S.; Argaves, W.S.; Pytela, R.; Arat, H.; Krusius, T.; Pierschbacher, M.D.; Ru  
 Proc. Natl. Acad. Sci. U.S.A. 83, 8614-8618, 1986  
 A:Title: cDNA and amino acid sequences of the cell adhesion protein receptor recognizing  
 A:Reference number: A26482; PMID:87041504; PMID:2430295  
 A:Accession: A26482  
 A:Molecule type: mRNA  
 A:Residues: 413-1048 <SUZ>  
 R:Cheresh, D.A.; Smith, J.W.; Cooper, H.M.; Quaranta, V.  
 Cell 57, 59-69, 1989  
 A:Title: A novel vitronectin receptor integrin (alpha-v beta-x) is responsible for dist  
 A:Reference number: A32287; PMID:89195223; PMID:2467745  
 A:Accession: A32287  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 31-35, 'X', '37-41 <CHE>  
 A:Experimental source: melanoma cell M21  
 A:Accession: B32287  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 31-33, 'X', '35, 'X', '37-41 <CH2>  
 A:Experimental source: lung carcinoma cell UCLA-P3  
 R:Iam, S.C.T.; Plow, E.F.; D'Souza, S.E.; Cheresh, D.A.; Freilinger III, A.L.; Ginsberg,  
 U. Biol. Chem. 264, 3742-3749, 1989  
 A:Title: Isolation and characterization of a platelet membrane protein related to the v  
 A:Reference number: A30298; PMID:89139425; PMID:2465293  
 A:Accession: A30298  
 A:Molecule type: protein  
 A:Residues: 31-35, 'X', '37-40 <LAM>  
 R:Smith, J.W.; Cheresh, D.A.  
 J. Biol. Chem. 265, 2168-2172, 1990  
 A:Title: Integrin (alpha-v beta-3)-ligand interaction. Identification of a heterodimeric  
 A:Reference number: A35035; PMID:90130470; PMID:1688848  
 A:Accession: A35035  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 66-72;169-171, 'X', 173-176;221-230;255-258, 'X', 260;325-328;342-351;466-473 <  
 C:Genetics:  
 A:Gene: GDB:ITGAV; VNRA  
 A:Cross-references: GDB:120491; OMIM:193210  
 A:Map position: 2q31-2q32  
 C:Superfamily: Integrin alpha-2b chain  
 C:Keywords: cell adhesion; duplication; glycoprotein; heterodimer; receptor; transmembr  
 F:1-30/Domain: signal sequence #status predicted <Sig>  
 F:31-1048/Product: integrin alpha-v chain #status experimental <M2>  
 F:31-992/Domain: extracellular #status predicted <EXT>  
 F:993-1016/Domain: transmembrane #status predicted <TM>  
 F:1017-1048/Domain: intracellular #status predicted <INT>

Query Match 43.9%; Score 50; DB 2; Length 1048;  
 Best Local Similarity 38.1%; Pred. No. 65;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db 1 KLGFFAKKIPPEEKREKLE 21  
 1017 RMGFFKVRPPQEEQEREQLQ 1037

RESULT 10  
 T12766  
 probable lipoprotein yolk - Bacillus subtilis phage SPBC2

C:Species: *Bacillus subtilis* phage SPB2  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 15-Oct-1999  
 C:Accession: T12766; F63907  
 R:Lazarevic, V.; Dueterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.  
 Submitted to the EMBL Data Library, August 1997  
 A:Description: The complete nucleotide sequence of the *Bacillus subtilis* SPB2act prophage  
 A:Reference number: 217583  
 A:Accession: T12766  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-236 <LAZ>  
 A:Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025480; PIDN:AC12975.1  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Aevedo, V.; Berth  
 C.: Bron, S.; Brouillet, C.V.; Brunsch, C.V.; Caldwell, B.; Capuano, M.; Carter, N.M.; Ch  
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabel, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall  
 lech, J.; Harwood, C.R.; Hentzel, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle  
 Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Saadate, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleif, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Serd  
 A.: Kench, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033; PMID:9354377  
 A:Accession: F63907  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-236 <KUN>  
 A:Cross-references: GB:Z59115; GB:AL009126; NID:g2634478; PIDN:CA14083.1; PID:el18312;  
 A:Experimental source: strain 168  
 A:Genetics:  
 A:Gene: yokB

Query Match 43.4%; Score 49.5; DB 2; Length 236;  
 Best Local Similarity 84.6%; Pred. No. 17;  
 Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 8 KKI-PEEKREK 19  
 ||:|||||  
 Db 205 KKVPEEKREK 217

RESULT 11  
 B60017  
 outer capsid protein VP2 - bluetongue virus (serotype 3, strain South Africa-VACC)  
 C:Species: bluetongue virus  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 16-Jul-1999  
 C:Accession: B60017  
 R:Gould, A.R.; Pritchard, L.I.  
 Virus Res. 17, 31-52, 1990  
 A:Title: Relationships amongst bluetongue viruses revealed by comparisons of capsid and  
 A:Reference number: A60017; MUID:91021485; PMID:2171239  
 A:Accession: B60017  
 A:Molecule type: genomic RNA  
 A:Residues: 1-959 <GOU>  
 A:Cross-references: GB:X55801; NID:g297130; PIDN:CAA39323.1; PID:g297131  
 C:Genetics:  
 A:Map position: segment 2  
 C:Superfamily: bluetongue virus VP2 protein  
 C:Keywords: capsid protein; glycoprotein  
 F:749; 910/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 43.4%; Score 49.5; DB 1; Length 959;  
 Best Local Similarity 57.1%; Pred. No. 69;  
 Matches 12; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 2 LGFFAHHKKIPEEKREKLEQ 22  
 |||:|||||  
 Db 791 LNFF-----PSYKREKLEK 806

RESULT 12  
 AC1466  
 B. subtilis YacI protein homolog lin0266 [imported] - *Listeria innocua* (strain Clp11262  
 C:Species: *Listeria innocua*  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C:Accession: AC1466  
 R:Glaser, P.; Frangoul, L.; Buchteler, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fain, H  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Krefel, T.; Kunz, M.; Kunst, F.; Kurapkac, G.; Magueno, E.; Maitounam, A.; Ma  
 ok, C.; Schlueter, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A:Title: Comparative genomics of *Listeria* species  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AC1466  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-357 <GLA>  
 A:Cross-references: GB:AL592022; PIDN:CA095499.1; PID:g16412695; GSPDB:GN00178  
 A:Experimental source: strain Clp11262  
 C:Genetics:  
 A:Superfamily: conserved hypothetical protein yacI

Query Match 43.0%; Score 49; DB 2; Length 357;  
 Best Local Similarity 71.4%; Pred. No. 30;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREK 19  
 |||:|||||  
 Db 146 AKKTPPEEKREK 159

RESULT 13  
 T06108  
 hypothetical protein TSJ17.190 - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
 C:Accession: T06108  
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.;  
 Submitted to the Protein Sequence Database, March 1999  
 A:Reference number: Z15184  
 A:Accession: T06108  
 A:Molecule type: DNA  
 A:Residues: 1-615 <BEV>  
 A:Cross-references: EMBL:AL035708; GSPDB:GN00062; ATSP:TSJ17.190  
 A:Experimental source: cultivar Columbia; BAC clone TSJ17  
 C:Genetics:  
 A:Gene: ATSP:TSJ17.190  
 A:Map position: 4

Query Match 43.0%; Score 49; DB 2; Length 615;  
 Best Local Similarity 71.4%; Pred. No. 52;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 KKIPEEKREKLE 21  
 |||:|||||  
 Db 396 KKIPEEKREKLE 409

RESULT 14  
 S74279  
 hypothetical protein YCL061c - yeast (*Saccharomyces cerevisiae*)  
 N:Alternate names: hypothetical protein YCL061c  
 C:Species: *Saccharomyces cerevisiae*  
 C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 29-Oct-1999  
 C:Accession: S74279; S19392; S19391; S29373; S21360  
 R:Voet, M.; Volckaert, G.  
 Submitted to the Protein Sequence Database, September 1996  
 A:Reference number: S74277  
 A:Accession: S74279

A/Molecule type: DNA  
 A/Residues: 1-853 <VOE>  
 A/Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA2405.1; PID:e309033; PID:g190711  
 A/Note: this is a revision to the sequence from reference S19391  
 R/Rasmussen, S.W.; von Wettstein, D.  
 submitted to the Protein Sequence Database, March 1992  
 A/Reference number: S19391  
 A/Accession: S19392  
 A/Molecule type: DNA  
 A/Residues: 'MKLPIMNHLVVKMTIVNPIVKRKMKL',346-619,'LP',621,'P',623-638,'LRTIALM' <RAS>  
 A/Cross-references: EMBL:X59720; MIPS:YCL061C  
 A/Note: this sequence has been revised in reference S74279  
 A/Note: this was assumed to be the complete sequence of protein YCL061C  
 A/Accession: S19391  
 A/Molecule type: DNA  
 A/Residues: 1-314,'IEW' <RAW>  
 A/Cross-references: EMBL:X59720  
 A/Note: this sequence has been revised in reference S74279  
 A/Note: this was assumed to be protein YCL060C  
 R/Kern, L.  
 Nucleic Acids Res. 18, 5279, 1990  
 A/Title: The URK1 gene of *Saccharomyces cerevisiae* encoding uridine kinase.  
 A/Reference number: S29373; MUID:90384830; PMID:2169608  
 A/Accession: S29373  
 A/Molecule type: DNA  
 A/Residues: 417-504,'V',506-564,566-596 <KER>  
 A/Cross-references: EMBL:X53998; NID:g4771; PIDN:CAA37945.1; PID:g4772  
 C/Genetics:  
 A/Map position: 3L  
 A/Note: YCL061C

Query Match 43.0%; Score 49; DB 2; Length 853;  
 Best Local Similarity 56.2%; Pred. No. 72;  
 Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 7 HKKIPPEKREKLEQ 22  
 Db 277 NKRIROKREKREKLE 292

## RESULT 15

C71956  
 Probable endopeptidase Clp ATP-binding chain - *Helicobacter pylori* (strain J99)  
 N/Alternate names: ATP-dependent Clp proteinase regulatory chain  
 N/Contents: adenosinetriphosphatase (EC 3.6.1.3)  
 C/Species: *Helicobacter pylori*  
 A/Variety: strain J99  
 C/Date: 12-Feb-1999 #sequence-revision 12-Feb-1999 #text\_change 19-Jan-2001  
 C/Accession: C71956  
 R/Alm, R.A.; Ling, L.S.L.; Molt, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
 ; Ives, C.; Gibson, R.; Menberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
 Nature 397, 176-180, 1999  
 A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
 A/Reference number: A71800; MUID:99120557; PMID:9923682  
 A/Accession: C71956  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-856 <ARN>  
 A/Cross-references: GB:AE001462; GB:AE001439; NID:g4154760; PIDN:AAD05825.1; PID:g415476  
 A/Experimental source: strain J99  
 C/Genetics:  
 A/Gene: clpB  
 C/Function:  
 A/Description: allows clpB to hydrolyze polypeptides and proteins, probably by a chaperon  
 e activity; ATP hydrolysis is required for Clp ATP-binding chain  
 C/Superfamily: endopeptidase Clp ATP-binding chain  
 C/Keywords: ATP; hydrolase; molecular chaperone; nucleotide binding; P-loop  
 F:204-211/Region: nucleotide-binding motif A (P-loop)  
 F:271-276/Region: nucleotide-binding motif B  
 F:605-612/Region: nucleotide-binding motif A (P-loop)  
 F:673-678/Region: nucleotide-binding motif B  
 F:210/Binding site: ATP (Lys) #status predicted  
 F:611/Binding site: ATP (Lys) #status predicted

Query Match 43.0%; Score 49; DB 2; Length 856;  
 Best Local Similarity 53.3%; Pred. No. 72;  
 Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
 Oy 7 HKKIPPEKREKLE 21  
 Db 500 YSKIPPEKREKLEQ 514

Search completed: July 16, 2003, 07:56:52  
 Job time: 3.11294 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:42 ; Search time 0.572433 Seconds

(without alignments)  
1594.036 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114  
Sequence: 1 KLGFFAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	1167	ITAG_HUMAN	O75578 homo sapien
2	51	44.7	128	YQ40_BACAN	Q91M22 bacillus an
3	51	44.7	856	CLPB_HELPY	P71404 helicobacte
4	50	43.9	357	Y234_LISMO	O48762 listeria mo
5	50	43.9	505	SPKD_SYNY3	P54735 synchocyst
6	50	43.9	1034	ITAV_CHICK	P26008 gallus gall
7	50	43.9	1044	ITAV_MOUSE	P43406 mus musculu
8	50	43.9	1048	ITAV_HUMAN	P06756 homo sapien
9	49.5	43.4	959	VP2_BT3V	O06998 bluetongue
10	49	43.0	357	Y266_LISIN	O92541 listeria in
11	49	43.0	853	YCGA_YEAST	P25588 saccharomyc
12	49	43.0	856	CLPB_HELPY	O92M11 helicobacte
13	48.5	42.5	407	M21_STRPY	P50468 streptococc
14	48	42.1	573	ERGI_RAT	P52020 rattus norv
15	48	42.1	1526	TP2A_RAT	P41516 rattus norv
16	47	41.2	112	FLIN_AOUAE	O67495 aquifex aeo
17	47	41.2	199	HMGA_MOUSE	O54879 mus musculu
18	47	41.2	692	YKOC_YEAST	P36062 saccharomyc
19	46	40.4	308	AMIP_STRPN	P18766 streptococc
20	46	40.4	450	YKHS_YEAST	P36083 saccharomyc
21	46	40.4	1407	TRHY_RABIT	P37709 oryctolagus
22	45	39.5	101	RL12_METTL	O52706 methanococc
23	45	39.5	147	RL9_THETN	O81667 thermomast
24	45	39.5	333	AOX2_SOYBN	O41265 glycine max
25	45	39.5	339	RLAO_PYRFU	O82378 pyrococcus
26	45	39.5	527	ZIM2_HUMAN	O9NZV7 homo sapien
27	45	39.5	685	TRPE_PAROL	O93429 paralichthy
28	45	39.5	1022	TPS3_YEAST	P38426 saccharomyc
29	45	39.5	785	YAPA_SCHPO	O09863 schizosacch
30	44	38.6	102	DMS4_TAGAN	O93224 agalactinis
31	44	38.6	93	S108_HUMAN	P05109 homo sapien
32	44	38.6	184	YMKC_BACSU	P45870 bacillus su
33	44	38.6	349	ARSA_METUA	O58542 methanococc

34	44	38.6	350	Y170_METUA	O57634 methanococc
35	44	38.6	450	FEM2_RAT	O9WV7 rattus norv
36	44	38.6	475	TPS1_PICAN	O94213 picchia angu
37	44	38.6	477	TPS1_AOUAE	O67499 aquifex aeo
38	44	38.6	478	TPS1_CANAL	O92410 candida alb
39	44	38.6	566	PEP_YEAST	P11491 saccharomyc
40	44	38.6	568	DISC_DROME	P23792 drosophila
41	44	38.6	1033	ITAB_MOUSE	O9QUM0 mus musculu
42	44	38.6	1121	MYR1_HUMAN	O01538 homo sapien
43	43.5	38.2	1863	BRCL_HUMAN	P38398 homo sapien
44	43	37.7	100	RR14_LOTUA	O9BHC3 lotus japon
45	43	37.7	100	RR14_TOBAC	P06371 nicotiana t

## ALIGNMENTS

RESULT 1  
ITAG\_HUMAN STANDARD; PRT; 1167 AA.  
AC O75578; Q9UH28; 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-10 precursor.  
GN ITGA10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Articular chondrocytes;  
RX MEDLINE=98352078; PubMed=9685391;  
RA Camper L., Hellman U., Lundgren-Akerlund E.,  
RT "Isolation, cloning, and sequence analysis of the integrin subunit  
RT alpha10, a beta1-associated collagen binding integrin expressed on  
RT chondrocytes."  
RL J. Biol. Chem. 273:20383-20389(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart, and Endothelial cells;  
RX MEDLINE=20169197; PubMed=10702680;  
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,  
RA Wang S.-X., Langley R., Kristensen G.W.;  
RT "The integrin alpha10 subunit: expression pattern, partial gene  
RT structure, and chromosomal localization".  
RL Cytogenet. Cell Genet. 87:238-244(1999).  
CC -1- FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10  
CC ASSOCIATES WITH BETA-1.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN  
CC MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.  
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INTEGR) IS A VMFA DOMAIN. INTEGRINS  
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 VMFA DOMAIN.  
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: AF074015; AAC31952.1; -;  
CC DR EMBL: AF112345; AAF21944.1; -;  
CC DR EMBL: AF172723; AAF61638.1; -;  
CC DR HSSP: P17301; IAOX.  
CC Genew; HGNC:6135; ITGA10.

```

RA  SEQUENCE FROM N.A.
RA  Okunaka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lanke G.,
RA  Kumano S., Menter D., Martinez Y., Svensson R., Tatum L.R.,
RA  Brown A.E., Jackson P.J.;
RA  Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC -----
CC  EMBL, AF188935; AAF13645.1;
DR  Hypothetical protein; Plasmid.
RW  SEQUENCE 128 AA; 14687 MW;  D54C69BAD5A0AFA CRC64;
SQ -----
Oy  Query Match 44.7%; Score 51; DB 1; Length 128;
Db  Best local similarity 50.0%; Pred. NO. 2.2;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Oy 2 LGFFAHKKIPEEEKREEK 19
Db 9 MGFFGNKGKPASEKKDEK 26
Oy :::::|::|::|::|
Db 9 MGFFGNKGKPASEKKDEK 26
-----
RESULT 3
CLPB_HELPY
ID CLPB_HELPY STANDARD; PRT; 856 AA.
AC P71404.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CLPB protein.
DE CLPB OR HP0264.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
[1]
RN RP SEQUENCE FROM N.A.
RC RA STRAIN=NTCC 11637;
RN RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN RP SEQUENCE FROM N.A.
RC RA STRAIN=26695 / ATCC 700392;
RN RL MEDLINE=97394467; PubMed=252185;
RX TX Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson P., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weyden J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC -----

```



CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AB046600; BAB17036.1; -  
DR EMBL; D64005; BAA10726.1; -  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR003646; SH3\_bac.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR Pfam; PF00069; Pkinase; 1.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00287; SH3D; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;  
KW ATP-binding; Complete proteome.  
FT DOMAIN 9 271 PROTEIN KINASE.  
FT NP\_BIND 15 23 ATP (BY SIMILARITY).  
FT BINDING 40 40 ATP (BY SIMILARITY).  
FT ACT\_SITE 136 136 BY SIMILARITY.  
SQ SEQUENCE 505 AA; 55213 MW; C4F12A1886C4D51C CRC64;

Query Match 43.9%; Score 50; DB 1; Length 505;  
Best Local Similarity 58.8%; Pred. No. 12;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 6 AHKKIPKREKREKLEQ 22  
Db 376 AEOKIAENKREKRELEQ 392

RESULT 6  
ITAV\_CHICK STANDARD; PRT; 1034 AA.  
ID ITAV\_CHICK  
AC P26008;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-V precursor (Vltroectin receptor alpha subunit).  
GN ITGAV.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
CX NCBI\_TaxID=9031;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=91104936; PubMed=1703004;  
RA Borey B., Reichardt L.F.;  
RT "Chick integrin alpha V subunit molecular analysis reveals high  
RT conservation of structural domains and association with multiple beta  
RT subunits in embryo fibroblasts."  
RL Biochemistry 29:10191-10198(1990).  
CC -1- FUNCTION: THE ALPHA-V INTEGRINS ARE RECEPTORS FOR VITRONECTIN,  
CC CYTOTACTIN, FIBRONECTIN, FIBRINOGEN, LAMININ, MATRIX  
CC METALLOPROTEINASE-2, OSTEOPOINTIN, PROTHROMBIN, THROMBOSPONDIN AND  
CC VON WILLEBRAND FACTOR. THEY RECOGNIZE THE SEQUENCE R-G-D IN A WIDE  
CC ARRAY OF LIGANDS.  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA  
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A  
CC DISULFIDE BOND. ALPHA-V ASSOCIATES WITH EITHER BETA-1, BETA-3,  
CC BETA-5, BETA-6 OR BETA-8.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M60517; AAA49138.1; -  
DR PIR; A36108; A36108.  
DR HSSP; P06756; 1JW2.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR Pfam; PF01839; FG-GAP; 5.  
DR PRINTS; PR01185; INTEGRIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR PROSITE; PS00242; INTEGRIN ALPHA; 1.  
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
KW Signal; Repeat; Calcium.  
FT SIGNAL 1 19  
FT CHAIN 20 1034  
FT CHAIN 20 875  
FT CHAIN 20 978  
FT DOMAIN 20 978  
FT TRANSMEM 979 1002  
FT DOMAIN 1003 1034  
FT REPEAT 35 96  
FT REPEAT 101 166  
FT REPEAT 167 224  
FT REPEAT 237 290  
FT REPEAT 291 350  
FT REPEAT 355 415  
FT REPEAT 420 473  
FT CA\_BIND 248 256  
FT CA\_BIND 302 310  
FT CA\_BIND 367 375  
FT CA\_BIND 431 439  
FT SITE 1005 1009  
FT DISULFID 77 85  
FT DISULFID 126 146  
FT DISULFID 160 173  
FT DISULFID 479 488  
FT DISULFID 494 551  
FT DISULFID 612 618  
FT DISULFID 684 697  
FT DISULFID 838 890  
FT DISULFID 895 900  
FT CARBOHYD 62 278  
FT CARBOHYD 284 284  
FT CARBOHYD 540 540  
FT CARBOHYD 601 601  
FT CARBOHYD 690 690  
FT CARBOHYD 821 821  
FT CARBOHYD 837 837  
FT CARBOHYD 860 860  
FT CARBOHYD 931 931  
FT CARBOHYD 951 951  
FT CARBOHYD 959 959  
FT CARBOHYD 966 966  
SQ SEQUENCE 1034 AA; 114388 MW; D76B08BA692DC684 CRC64;

Query Match 43.9%; Score 50; DB 1; Length 1034;  
Best Local Similarity 38.1%; Pred. No. 26;  
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 1 KLGFPAHKIKPEEKREKLE 21  
Db 1003 RMGFFKVRPPQREOREBOLQ 1023

RESULT 7  
ITAV\_MOUSE STANDARD; PRT; 1044 AA.  
ID ITAV\_MOUSE  
AC P43406;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)

FT	CA_BIND	443	451	POTENTIAL.
FT	SITE	1015	1019	OFFER MOTIF.
FT	DISULFID	89	97	BY SIMILARITY.
FT	DISULFID	138	158	BY SIMILARITY.
FT	DISULFID	172	185	BY SIMILARITY.
FT	DISULFID	491	502	BY SIMILARITY.
FT	DISULFID	508	565	BY SIMILARITY.
FT	DISULFID	626	632	BY SIMILARITY.
FT	DISULFID	698	711	BY SIMILARITY.
FT	DISULFID	852	900	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	905	910	BY SIMILARITY.
FT	CARBOHYD	74	74	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	290	290	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	296	296	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	704	704	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	835	835	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	851	851	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	869	869	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	941	941	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	969	969	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	976	976	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	1044 AA;	115277 MW;	F06A8FF22705183D CRC64;

Query Match	Best Local Similarity	43.9%;	Score 50;	DB 1;	Length 1044;
Matches	8;	Conservative 1%;	Pred. No. 26;	Mismatches 7;	Indels 6;
					Gaps 0;

Qy	1	KLGFEAHKKIPEEKREKXLE 21
Db	1013	RMGFPRVPRPQEQERQLO 1033

RESULT 8			
ID	ITAV_HUMAN	STANDARD;	PRT; 1048 AA.
AC	P06756;		
DT	01-JUN-1988	(Rel. 06, Created)	
DT	01-AUG-1988	(Rel. 08, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Integrin alpha-V precursor (vitronectin receptor alpha subunit) (CD51 antigen).		
GN	ITGAV OR VNRA.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=68007656; PubMed=2443500;		
RA	Suzuki S., Argreaves W.S., Arai H., Languino L.R., Pierschbacher M.D.,		
RA	Ruoslahti E.,		
RT	"Amino acid sequence of the vitronectin receptor alpha subunit and		
RT	comparative expression of adhesion receptor mRNAs."		
RL	J. Biol. Chem. 262:14080-14085(1987).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20422500; PubMed=10965141;		
RA	Stam M.A., Field S., Barnes M.R., Shaikh N., Ellington K.,		
RA	Murphy K.E., Spurr N.K., Campbell D.A.,		
RT	"Cloning and characterisation of ITGAV, the genomic sequence for human		
RT	cell adhesion protein (vitronectin) receptor alpha subunit, CD51."		
RL	Cytogene. Cell Genet. 89:268-271(2000).		
RN	[3]		
RP	SEQUENCE OF 413-1048 FROM N.A.		
RX	MEDLINE=87041504; PubMed=2430295;		
RA	Suzuki S., Argreaves W.S., Pyrela R., Arai H., Krusius T.,		
RA	Pierschbacher M.D., Ruoslahti E.,		
RT	"CDNA and amino acid sequences of the cell adhesion protein receptor		
RT	recognizing vitronectin reveals a transmembrane domain and homologues		
RT	with other adhesion protein receptors."		
RL	Proc. Natl. Acad. Sci. U.S.A. 83:8614-8618(1986).		
RN	[4]		

RP SEQUENCE OF 31-41.  
 RX MEDLINE=89195223; PubMed=2467745;  
 RA Chershe D.A., Smith J.W., Cooper H.M., Quaranta V.;  
 RT "A novel vitronectin receptor integrin (alpha v beta x) is  
 RT responsible for distinct adhesive properties of carcinoma cells.";  
 RT Cell 57:59-69(1989).  
 RL [5]  
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 31-987.  
 RX MEDLINE=21482770; PubMed=11546839;  
 RA Xiong J.P., Stehle T., Dieffenbach B., Zhang R., Dunker R., Scott D.L.,  
 RA Joachimiak A., Goodman S.L., Arnaout M.A.;  
 RT "Crystal structure of the extracellular segment of integrin alpha  
 RT vbeta3.";  
 RT Science 294:339-345(2001).  
 RL -!- FUNCTION: The alpha-V integrins are receptors for vitronectin,  
 CC cytoactin, fibronectin, fibrinogen, laminin, matrix  
 CC metalloproteinase-2, osteopontin, osteomodulin, prothrombin,  
 CC thrombospondin and von Willebrand factor. They recognize the  
 CC sequence R-G-D in a wide array of ligands.  
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA  
 CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A  
 CC DISULFIDE BOND. ALPHA-V ASSOCIATES WITH EITHER BETA-1, BETA-3,  
 CC BETA-5, BETA-6 OR BETA-8 SUBUNIT.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -!- DATABASE: NAME=PROV; NOTE=CD guide CDS1 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cds1.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M14648; AAA36808.1; -;  
 DR EMBL; AF251841; AAG03000.1; JOINED.  
 DR EMBL; AF251818; AAG03000.1; JOINED.  
 DR EMBL; AF251819; AAG03000.1; JOINED.  
 DR EMBL; AF251820; AAG03000.1; JOINED.  
 DR EMBL; AF251821; AAG03000.1; JOINED.  
 DR EMBL; AF251822; AAG03000.1; JOINED.  
 DR EMBL; AF251823; AAG03000.1; JOINED.  
 DR EMBL; AF251824; AAG03000.1; JOINED.  
 DR EMBL; AF251825; AAG03000.1; JOINED.  
 DR EMBL; AF251826; AAG03000.1; JOINED.  
 DR EMBL; AF251827; AAG03000.1; JOINED.  
 DR EMBL; AF251828; AAG03000.1; JOINED.  
 DR EMBL; AF251829; AAG03000.1; JOINED.  
 DR EMBL; AF251830; AAG03000.1; JOINED.  
 DR EMBL; AF251831; AAG03000.1; JOINED.  
 DR EMBL; AF251832; AAG03000.1; JOINED.  
 DR EMBL; AF251833; AAG03000.1; JOINED.  
 DR EMBL; AF251834; AAG03000.1; JOINED.  
 DR EMBL; AF251835; AAG03000.1; JOINED.  
 DR EMBL; AF251836; AAG03000.1; JOINED.  
 DR EMBL; AF251837; AAG03000.1; JOINED.  
 DR EMBL; AF251838; AAG03000.1; JOINED.  
 DR EMBL; AF251839; AAG03000.1; JOINED.  
 DR EMBL; AF251840; AAG03000.1; JOINED.  
 DR PIR; A27421; A27421.  
 DR PIR; A26482; A26482.  
 DR PIR; A32287; A32287.  
 DR PIR; B32287; B32287.  
 DR PDB; 1UV2; 17-OCT-01.  
 DR GeneW; HGNC:6150; ITGAV.  
 DR MIM; 193210; -;  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR Pfam; PF00357; Integrin\_A\_1.  
 DR Pfam; PF01839; FG-GAP\_5.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR SMART; SMO0191; int\_alpha; 5.

DR PROSITE; PS00242; INTEGRIN ALPHA; 1.  
 KW Integrin; Cell adhesion; Glycoprotein; Transmembrane; Signal;  
 KM Repeat; Calcium; 3D-structure.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1048 INTEGRIN ALPHA-V.  
 FT CHAIN 31 889 INTEGRIN ALPHA-V HEAVY CHAIN.  
 FT CHAIN 891 1048 INTEGRIN ALPHA-V LIGHT CHAIN.  
 FT DOMAIN 31 992 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 993 1016 POTENTIAL.  
 FT DOMAIN 1017 1048 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 46 108 FG-GAP 1.  
 FT REPEAT 113 178 FG-GAP 2.  
 FT REPEAT 179 236 FG-GAP 3.  
 FT REPEAT 249 302 FG-GAP 4.  
 FT REPEAT 303 362 FG-GAP 5.  
 FT REPEAT 367 427 FG-GAP 6.  
 FT REPEAT 432 483 FG-GAP 7.  
 FT CA BIND 260 268 POTENTIAL.  
 FT CA BIND 314 322 POTENTIAL.  
 FT CA BIND 379 387 POTENTIAL.  
 FT CA BIND 443 451 POTENTIAL.  
 FT SITE 1019 1023 GPEKR MOTIF.  
 FT DISULFID 89 97 BY SIMILARITY.  
 FT DISULFID 138 158 BY SIMILARITY.  
 FT DISULFID 172 185 BY SIMILARITY.  
 FT DISULFID 491 502 BY SIMILARITY.  
 FT DISULFID 508 565 BY SIMILARITY.  
 FT DISULFID 626 632 BY SIMILARITY.  
 FT DISULFID 698 711 BY SIMILARITY.  
 FT DISULFID 852 904 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 909 914 BY SIMILARITY.  
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 704 704 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 973 973 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 425 425 W -> R (IN REF. 2).  
 FT CONFLICT 783 783 I -> V (IN REF. 2).  
 FT CONFLICT 1039 1039 H -> R (IN REF. 2).  
 SQ SEQUENCE 1048 AA; 116051 MW; 3648B0B350D6D07 CRC64;  
 Query Match 43.9%; Score 50; DB 1; Length 1048;  
 Best Local Similarity 38.1%; Pred. No. 26;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;  
 Oy 1 KLGFPAAHKIPPEEKREKLE 21  
 Db 1017 RMGFFKVRPPQREOERQLO 1037  
 RESULT 9  
 ID\_VP2\_BT3V STANDARD; PRT; 959 AA.  
 AC 006938;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Outer capsid protein VP2.  
 GN S2.  
 OS Bluecough virus (serotype 3 / isolate South Africa-vaccine).  
 OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.  
 NX NCBI\_TaxID=36424;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91021485; PubMed=2171239;

RA Gould A.R., Prichard L.I.;  
 RT "Relationships amongst bluetongue viruses revealed by comparisons of  
 RT capsid and outer coat protein nucleotide sequences.";   
 RL Virus Res. 17:31-52(1990).  
 CC -1- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VPS)  
 CC WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE  
 CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.  
 CC -1- SIMILARITY: BELONGS TO THE REOVIRUSES VP2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X55801; CAA39323.1; -;  
 DR PIR; B60017; B60017;  
 DR InterPro; IPR001742; Orbi\_VP2.  
 DR Pfam; PF00898; Orbi\_VP2; 1.  
 DR Prodom; PD002938; Orbi\_VP2; 1.  
 KW Coat protein.  
 SQ SEQUENCE 959 AA; 112163 MW; 41B230B31803588F CRC64;  
 Query Match 43.4%; Score 49.5; DB 1; Length 959;  
 Best Local Similarity 57.1%; Pred. No. 28;  
 Matches 12; Conservative 1; Mismatches 3; Indels 5; Gaps 1;  
 QY 2 LGFPAHKKIPBEERKELEQ 22  
 Db 791 LNFF-----PSYERKEELEER 806  
 RESULT 10  
 Y266\_LISIN STANDARD; PRT; 357 AA.  
 ID Y266\_LISIN  
 AC Q92F41;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein Lin0266.  
 GN Lin0266.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 CC NCBI\_TaxID=1642;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / Serovar 6a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaeser P., Francaul L., Buchrieser C., Ruenick C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chevrouat F., Couve E., de Daruvar A., Denoux F.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauguet O.,  
 RA Eutlier K.-D., Fehli H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapat G.,  
 RA Medueno E., Maitounem A., Mata Vicente J., Ng E., Medjari H.,  
 RA Nordstedt G., Novella S., de Padlos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";   
 RL Science 294:849-852(2001).  
 CC -1- SIMILARITY: BELONGS TO THE YCF81 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AL596164; CAC95499.1; -;  
 DR Listlist; LIN00266; -;  
 DR InterPro; IPR002792; TRAM.  
 DR Pfam; PF01938; TRAM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 357 AA; 39433 MW; 23A6C248979167F4 CRC64;  
 Query Match 43.0%; Score 49; DB 1; Length 357;  
 Best Local Similarity 71.4%; Pred. No. 12;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 6 AHKKIPBEERKEEK 19  
 Db 146 AKKTPBEERKEEK 159  
 RESULT 11  
 YCG1\_YEAST STANDARD; PRT; 853 AA.  
 ID YCG1\_YEAST  
 AC P25588; P25589; P27513; P87003;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Hypothetical 97.9 kDa protein in CHAI-KRI intergenic region.  
 GN YCL061C OR YCL61C/YCL60C.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes.  
 CC NCBI\_TaxID=4932;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Rasmussen S.W., von Wettstein D.;  
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP REVISIONS.  
 RA Gromadka R.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 384-596 FROM N.A.  
 RC STRAIN=ATCC 28383 / FL100;  
 RX MEDLINE=90384830; PubMed=2169608;  
 RA Kern L.;  
 RT "The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.";   
 RL Nucleic Acids Res. 18:5279-5279(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59720; CAA42405.1; -;  
 DR EMBL; X53998; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; S19391; S19391.  
 DR PIR; S19392; S19392.  
 DR PIR; S21360; S21360.  
 DR PIR; S29373; S29373.  
 DR SGD; S0000566; YCL061C.  
 KW Hypothetical protein.  
 FT CONFLICT 505 505  
 FT CONFLICT 567 567  
 SQ SEQUENCE 853 AA; 97946 MW; 16E09FCC0B246D1 CRC64;  
 Query Match 43.0%; Score 49; DB 1; Length 853;  
 Best Local Similarity 56.2%; Pred. No. 29;  
 Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 QY 7 HKKIPBEERKEELEQ 22  
 Db 277 NKRIROKREKREKLE 292

```

RESULT 12
CLPB_HELPJ STANDARD: PRT: 856 AA.
ID CLPB_HELPJ
AC Q92MH1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CLPB protein
GN CLPB OR JH0249.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_Taxid=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carnel G.,
RA Tummino P.J., Carnuso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -1 SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
CC ATP-DEPENDENT PROTEASE (BY SIMILARITY)
CC -1 SIMILARITY: BELONGS TO THE CLPB/CLPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001462; AAD05825.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR001270; Chaperin_in_clpb/B.
DR InterPro; IPR004176; Clp_N.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02861; Clp_N; 2.
DR PRINTS; PR00300; CLPBPROTEASEA.
DR PRODOM; PR000739; GSPIT_E; 1.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPBAB_1; 1.
DR PROSITE; PS00871; CLPBAB_2; 1.
KW Chaperone; ATP-binding; Repeat; Complete proteome.
FT DOMAIN 159 406
FT NP_BIND 531 722
FT NP_BIND 204 211
FT NP_BIND 605 612
FT NP_BIND 605 612
SQ SEQUENCE 856 AA; 96730 MW; 0B6658F6946588A CRC64;

Query Match 43.0%; Score 49; DB 1; Length 856;
Best Local Similarity 53.3%; Pred. No. 29;
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

```

```

GN EMBL2.1.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=T2/44/RB4/119;
RX MEDLINE=92104662; PubMed=1370269;
RA Bessen D.E., Fischetti V.A.;
RT "Nucleotide sequences of two adjacent M or M-like protein genes of
RT group A streptococci: different RNA transcript levels and
RT identification of a unique immunoglobulin A-binding protein...";
RL Infect. Immun. 60:124-135(1992).
RN [2]
RP SEQUENCE OF 138-305 FROM N.A.
RX STRAIN=T2/44/RB4;
RX MEDLINE=91079780; PubMed=2258705;
RA Bessen D.E., Fischetti V.A.;
RT "Differentiation between two biologically distinct classes of group A
RT streptococci by limited substitutions of amino acids within the
RT shared region of M protein-like molecules.";
RL J. Exp. Med. 172:1757-1764(1990).
CC -1 FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC -1 SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -1 SIMILARITY: TO OTHER M PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61276; CAA43581.1; -
DR EMBL; X56398; CAA38808.1; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02370; M; 9.
DR PRINTS; PR00015; GPOSANCHOR.
DR TIGRfams; TIGR01167; LPXTG_anchor; 1.
DR TIGRfams; TIGR01168; YSIRK_signal; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Virulence; Phagocytosis; Cell wall; Peptidoglycan-anchor; Repeat;
KW Antigen; Coiled coil; Signal.
FT SIGNAL 1 41
FT CHAIN 42 377
FT PROPEP 378 407
FT DOMAIN 81 94
FT REPEAT 81 87
FT REPEAT 88 94
FT DOMAIN 158 285
FT REPEAT 158 180
FT REPEAT 193 215
FT REPEAT 228 250
FT REPEAT 263 285
FT DOMAIN 116 157
FT DOMAIN 348 373
FT SITE 374 378
FT MOD_RES 377 377
SQ SEQUENCE 407 AA; 46466 MW; 33CA053B7DB3C1EA CRC64;

Query Match 42.5%; Score 48.5; DB 1; Length 407;
Best Local Similarity 58.8%; Pred. No. 16;
Matches 10; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

```



OY 7 HKKPEEKRE-EKLEQ 22  
 DB 85 HKKVEEHKKDHETLEK 101

## RESULT 14

ERG1\_RAT STANDARD; PRT; 573 AA.  
 AC P52020;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Squalene monooxygenase (EC 1.14.99.7) (Squalene epoxidase) (SE).  
 GN SQUAL OR ERG1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=95113819; PubMed=7814369;  
 RA Sakakibara J., Matanabe R., Kanai Y., Ono T.;  
 RT "Molecular cloning and expression of rat squalene epoxidase.";  
 RL J. Biol. Chem. 270:17-20(1995).  
 CC -1- FUNCTION: CATALYZES THE FIRST OXYGENATION STEP IN STEROL  
 BIOSYNTHESIS AND IS SUGGESTED TO BE ONE OF THE RATE-LIMITING  
 ENZYMES IN THIS PATHWAY.  
 CC -1- CATALYTIC ACTIVITY: Squalene + AH(2) + O(2) = (S)-squalene-2,3-  
 epoxide + A + H(2)O.  
 CC -1- COFACTOR: FAD.  
 CC -1- SUBUNIT: MAY FORM A COMPLEX WITH SQUALENE SYNTHASE.  
 CC -1- SUBCELLULAR LOCATION: Microsomal.  
 CC -1- SIMILARITY: BELONGS TO THE SQUALENE MONOOXYGENASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D37920; BAA07141.1; -  
 DR InterPro: IPR000733; Flav monooxygenase.  
 DR InterPro: IPR000205; NAD binding.  
 DR Pfam: PF01360; Monooxygenase; 1.  
 DR Oxidoreductase; Flavoprotein; FAD; Transmembrane.  
 FT TRANSMEM 20 40 POTENTIAL.  
 FT TRANSMEM 60 80 POTENTIAL.  
 FT TRANSMEM 118 138 POTENTIAL.  
 FT TRANSMEM 545 565 POTENTIAL.  
 FT NP\_BIND 125 152 FAD (ADP PART) (POTENTIAL).  
 FT NP\_BIND 152 152 FAD (ADP PART) (POTENTIAL).  
 SQ SEQUENCE 573 AA; 64024 MW; 55222C911E2B8777 CRC64;

Query Match 42.1%; Score 48; DB 1; Length 573;  
 Best Local Similarity 52.9%; Pred. No. 27;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 LGFFAHKKPEEKRE 18  
 DB 75 IGFWAKSPSEKKEQ 91

## RESULT 15

TP2A\_RAT STANDARD; PRT; 1526 AA.  
 AC P41516;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA topoisomerase II, alpha isozyme (EC 5.99.1.3).  
 GN TOP2A OR TOP2 OR TOP2-2.

OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Testis;  
 RX MEDLINE=93290677; PubMed=8390253;  
 RA Park S.H., Yoon J.H., Kwon Y.D., Park S.D.;  
 RT "Nucleotide sequence analysis of the cDNA for rat DNA topoisomerase  
 II";  
 RL Biochem. Biophys. Res. Commun. 193:787-793(1993).  
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT  
 BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II  
 MAKES DOUBLE-STRAND BREAKS.  
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
 CC of double-stranded DNA.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH  
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES  
 CC RELAX ONLY NEGATIVE SUPERCOILS.  
 CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z46372; CAA66496.1; -  
 DR EMBL: Z19552; CAA79611.1; -  
 DR HSP: Z28676; -, NOT\_ANNOTATED\_CDS.  
 DR HSP: P06786; IBGW.  
 DR InterPro: IPR003594; ATPbind ATPase.  
 DR InterPro: IPR003957; CBFA NFYB topis.  
 DR InterPro: IPR001241; DNA Topoisom.  
 DR InterPro: IPR002205; DNA\_topoisom.  
 DR Pfam: PF00204; DNA\_gyraseB; 1.  
 DR Pfam: PF00521; DNA\_topoisom; 1.  
 DR Pfam: PF02518; HATPase\_c; 1.  
 DR PRINTS: PR00615; CCAATSUBUNTA.  
 DR PRODOM: PD000616; DNA\_topoisom; 1.  
 DR PRODOM: PD000742; DNA\_topoisom; 1.  
 DR SMART: SM00433; TOP2c; 1.  
 DR SMART: SM00434; TOP4c; 1.  
 DR PROSITE: PS00177; TOPOISOMERASE II; 1.  
 KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.  
 FT NP\_BIND 153 164 ATP (POTENTIAL).  
 FT ACT\_SITE 803 803 DNA CLEAVAGE (BY SIMILARITY).  
 SQ SEQUENCE 1526 AA; 173220 MW; A1961ABDB1B050F CRC64;

Query Match 42.1%; Score 48; DB 1; Length 1526;  
 Best Local Similarity 56.2%; Pred. No. 72;  
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 6 AHKKPEEKREKLE 21  
 DB 1091 AQQKVPDEENENEE 1106

Search completed: July 16, 2003, 07:52:53  
 Job time: 2.57243 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:02 ; Search time 2.25879 Seconds  
(without alignments)  
2006.842 Million cell updates/sec

Title: US-09-647-544-7  
Perfect score: 114  
Sequence: 1 XLGFFAHKKIPEERKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	46.5	214	16	08ZJVO
2	53	46.5	214	16	08ZJVO
3	53	46.5	1033	13	042598
4	52	45.6	1021	5	09GSGF4
5	51	44.7	180	17	08TJZ6
6	51	44.7	672	17	028247
7	51	44.7	1045	13	091292
8	51	44.7	1957	11	091YC9
9	50	43.9	403	5	09XYF8
10	50	43.9	1007	6	09GK48
11	50	43.9	1047	6	09MZD6
12	49.5	43.4	236	9	064016
13	49.5	43.4	236	16	032005
14	49.5	43.4	635	4	096LH8
15	49.5	43.4	1372	4	09HAW1
16	49.5	43.4	1388	4	09HAW2

17	49.5	43.4	2187	4	09H197	09H197 homo sapien
18	49.5	43.4	2254	4	09HCY0	09HCY0 homo sapien
19	49	43.0	615	10	03SMO3	09GK3 arabidopsis
20	49	43.0	2527	5	095W83	095W83 plasmodium
21	48.5	42.5	211	2	09S4B2	09S4B2 streptococ
22	48	42.1	401	5	09U9N9	09U9N9 trypanosoma
23	48	42.1	421	17	097VL7	097VL7 sulfolobus
24	48	42.1	936	12	08QRY7	08QRY7 chimpanzee
25	48	42.1	958	10	09AVP6	09AVP6 vicia faba
26	48	42.1	1034	6	09TUN4	09TUN4 oryctolagus
27	48	42.1	1070	12	08QL26	08QL26 sulfolobus
28	48	42.1	1464	5	096782	096782 drosophila
29	48	42.1	1464	5	09VX75	09VX75 drosophila
30	48	42.1	3487	2	09FPU0	09FPU0 microcytis
31	48	42.1	3487	2	09RMB3	09RMB3 microcytis
32	47.5	41.7	2343	6	018806	018806 canis faml
33	47	41.2	102	17	09UZV3	09UZV3 pyrococcus
34	47	41.2	208	5	09GYK4	09GYK4 caenorhabd
35	47	41.2	306	10	09SRV7	09SRV7 arabidopsis
36	47	41.2	309	4	09H4K1	09H4K1 homo sapien
37	47	41.2	512	16	08UDL7	08UDL7 agrobacteri
38	47	41.2	702	4	09NVC7	09NVC7 homo sapien
39	47	41.2	702	4	09H8Y1	09H8Y1 homo sapien
40	47	41.2	861	10	080738	080738 arabidopsis
41	46.5	40.8	347	17	096Z70	096Z70 sulfolobus
42	46	40.4	170	9	09ZXJ6	09ZXJ6 bacterioph
43	46	40.4	175	9	09MCM9	09MCM9 streptococ
44	46	40.4	278	10	09S186	09S186 arabidopsis
45	46	40.4	292	5	094824	094824 tetrahymena

## ALIGNMENTS

RESULT 1	08ZJVO	PRELIMINARY;	PRT;	214 AA.
AC	08ZJVO;			
DT	01-MAR-2002 (T-REMBLrel. 20, Created)			
DT	01-MAR-2002 (T-REMBLrel. 20, Last sequence update)			
DT	01-MAR-2002 (T-REMBLrel. 20, Last annotation update)			
DE	Membrane protein, transcribed divergently from serb.			
GN	SMP OR STM4577.			
OS	Salmonella typhimurium.			
OC	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Salmonella.			
OX	NCBI_TaxID=602;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtnay L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2."			
RT	Nature 413:852-856(2001).			
DR	EMBL; A5008915; AAL23392.1; -.			
KW	Complete proteome.			
SO	SEQUENCE 214 AA; 24274 MW; BBBD0C37C40A598 CRC64;			
Query Match	46.5%; Score 53; DB 16; Length 214;			
Best Local Similarity	71.4%; Pred. No. 8.3;			
Matches 10; Conservative	2; Mismatches 2; Indels 0; Gaps 0;			
Oy	6 AHKKIPEERKREK 19			
Db	200 ASKVPPEERKREK 213			
RESULT 2				

```

082079
ID 082079 PRELIMINARY; PRT; 214 AA.
AC 082079
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative membrane protein.
DE SMP OR STY4924.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OC NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jager K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AF677284; CAD03408.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 214 AA; 24279 MW; 2683D0DF9174BF80 CRC64;

Query Match 46.5%; Score 53; DB 16; Length 214;
Best Local Similarity 71.4%; Pred. No. 8.3;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREK 19
Db 200 ASKPEPEEREK 213

RESULT 3
ID 042598 PRELIMINARY; PRT; 1033 AA.
AC 042598;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Integrin alpha subunit.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodidae; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98211587; PubMed=9551862;
RA Joes T.O., Reintsch W.E., Brinker A., Klein C., Hausen P.;
RT "Cloning of the Xenopus integrin alpha(v) subunit and analysis of its
RT distribution during early development."
RL Int. J. Dev. Biol. 42:171-179(1998).
DR EMBL; U92006; AAB62090.1; -
DR HSP; P06756; IUV2.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PF00357; Integrin_A; 1.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
SQ SEQUENCE 1033 AA; 114025 MW; 35F73B0765C06BD6 CRC64;

Query Match 46.5%; Score 53; DB 13; Length 1033;
Best Local Similarity 47.6%; Pred. No. 37;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 1 KLGFAHKIPEEKREK 21
Db 1002 KLGFFKVRPPQETTEREQ 1022

RESULT 4
ID 09GSF4 PRELIMINARY; PRT; 1021 AA.
AC 09GSF4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Integrin alpha chain.
DE INTA.
OS Podocoryne carnea.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;
OC Hydrozoa; Podocoryne.
OC NCBI_TaxID=6096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21375967; PubMed=11482899;
RA Reber-Muller S., Studer R., Muller P., Yanze N., Schmid V.;
RT "Integrin and talin in the jellyfish Podocoryne carnea."
RL Cell Biol. Int. 25:753-769(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Reber-Muller S., Studer R., Muller P., Yanze N., Schmid V.;
RX Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF308651; AAG25993.1; -
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PRO1185; INTEGRIN.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN 1.
SQ SEQUENCE 1021 AA; 110967 MW; 28A8E782FF616DB2 CRC64;

Query Match 45.6%; Score 52; DB 5; Length 1021;
Best Local Similarity 55.6%; Pred. No. 50;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KLGFAHKIPEEKREK 18
Db 996 KKGFFKXKXGDEDEEE 1013

RESULT 5
ID 08TZ06 PRELIMINARY; PRT; 180 AA.
AC 08TZ06;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein PF1995.
GN PF1995.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OC NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB010292; AAB82119.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 180 AA; 21325 MW; 33BAE26104B200AE CRC64;

Query Match 44.7%; Score 51; DB 17; Length 180;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

QY 2 LGFFFAKKIPEEKREK 16  
DB 122 VGFFIHRKINEEBAK 136

RESULT 6  
ID 028247 PRELIMINARY; PRT; 672 AA.

AC 028247;  
DT 01-JAN-1998 (TEMBLrel. 05, Created)  
DT 01-JUN-1998 (TEMBLrel. 05, Last sequence update)  
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE Signal-transducing histidine kinase, putative.  
GN AF2032.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kellavag A.R., Graham D.E., Kyriades N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Cotton M.D., Spriggs T., Arlanch P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-  
reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 350:364-370(1997).  
DR EMBL; AE000963; AAB89224.1; -.  
DR TIGR; AF2032; -.

DR InterPro; IPR003661; His\_kinA.  
DR InterPro; IPR004359; HIS\_KIN\_sig.  
DR InterPro; IPR001610; PAC.  
DR InterPro; IPR007000; PAS-aseC.  
DR InterPro; IPR000014; PAS\_domain.  
DR Pfam; PF00785; PAC; 3.  
DR Pfam; PF00989; PAS; 4.  
DR Pfam; PF00512; signal; 1.  
DR SMART; SM00388; HiskA; 1.  
DR SMART; SM00086; PAC; 3.  
DR SMART; SM00091; PAS; 4.  
DR TIGRFAMs; TIGR00229; sensory\_box; 4.  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 672 AA; 78620 MW; F95DC675BDB8F749 CRC64;

Query Match 44.7%; Score 51; DB 17; Length 672;  
Best Local Similarity 55.6%; Pred. No. 47;  
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 FAHKKIPEEKREKLEQ 22  
DB 572 FAFKMEDEERRELLKQ 589

RESULT 7

Q91292 PRELIMINARY; PRT; 1045 AA.

AC Q91292;  
DT 01-NOV-1996 (TEMBLrel. 01, Created)  
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE Integrin.  
OS Pleurodeles waltlil (Iberian ribbed newt).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;

OC Pleurodeles.  
OX NCBI\_TaxID=8319;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95377519; PubMed=7649360;  
RA Alfandari D., Whitaker C.A., Desimone D.W., Darrivere T.,  
RT "Integrin alpha v subunit is expressed on mesodermal cell surfaces  
during amphibian gastrulation.";  
RL Dev. Biol. 170:249-261(1995).  
DR EMBL; X81108; CAAS7014.1; -.  
DR HSSP; P06756; IJVV2.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR Pfam; PF01839; FG-GAP; 5.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR SMART; SM00191; Int\_alpha; 5.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; UNKNOWN\_1.  
KW Integrin.  
SQ SEQUENCE 1045 AA; 115421 MW; 4DB34B766B0C648E CRC64;

Query Match 44.7%; Score 51; DB 13; Length 1045;  
Best Local Similarity 42.9%; Pred. No. 71;  
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 KLGFFAKKIPEEKREKLE 21  
DB 1014 KFGFFKRVBPQEGEREDLQ 1034

RESULT 8

Q91Y9C PRELIMINARY; PRT; 1957 AA.

AC Q91Y9C;  
DT 01-DEC-2001 (TEMBLrel. 19, Created)  
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE 53BP1 protein.  
GN TRP53BP1 OR 53BP1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB C;  
RA Adachi Y.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB C;  
RA Jullien D.;  
RT "kinetochore localisation of the DNA damage response component 53BP1  
during mitosis.";  
RL J. Cell Sci. 0:0-0(0).  
DR EMBL; AJ414734; CAC94013.1; -.  
DR MGD; MG11351320; Trp53bp1.  
DR InterPro; IPR001357; BRCT.  
DR InterPro; IPR002114; HPr\_Serp\_site.  
DR Pfam; PF00533; BRCT; 2.  
DR PROSITE; PS00172; BRCT; 2.  
DR PROSITE; PS00589; PTS\_HPR\_SER; UNKNOWN\_1.  
SQ SEQUENCE 1957 AA; 211438 MW; 9199C7F4B627A8F CRC64;

Query Match 44.7%; Score 51; DB 11; Length 1957;  
Best Local Similarity 69.2%; Pred. No. 1.3e+02;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 10 IPEEKREKLEQ 22  
DB 129 LPPEEKREELER 141

RESULT 9

09xyf8 PRELIMINARY; PRT; 403 AA.  
 ID 09xyf8  
 AC 09xyf8  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE Calreticulin.  
 GN CLR.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TUHUEN 2;  
 RX MEDLINE=99250150; PubMed=10233151;  
 RA Labriola C., Gazzulo J.J., Parodi A.J.;  
 RT "Trypanosoma cruzi calreticulin is a lectin that binds  
 RT monoglucosylated oligosaccharides but not protein moieties of  
 RT glycoproteins."  
 RL Mol. Biol. Cell. 10:1381-1394(1999).  
 RL EMBL: AF107115; AAD22175.1; -  
 DR InterPro: IPR001580; Calreticulin.  
 DR Pfam: PF00262; Calreticulin; 1.  
 DR PRINTS: PR00626; CALRETICULIN.  
 DR ProDom: PD001866; Calreticulin; 1.  
 DR PROSITE: PS00803; CALRETICULIN\_1; 1.  
 DR PROSITE: PS00804; CALRETICULIN\_2; UNKNOWN\_1.  
 DR PROSITE: PS00805; CALRETICULIN\_REPEAT; 1.  
 SQ SEQUENCE 403 AA; 46758 MW; AYBD560291AD071F CRC64;

Query Match 43.9%; Score 50; DB 5; Length 403;  
 Best Local Similarity 66.7%; Pred. No. 40;  
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 8 KKIPEEKREKLEHQ 22  
 Db 380 KENPEKDEBELE 394

RESULT 10  
 09GK48  
 ID 09GK48 PRELIMINARY; PRT; 1007 AA.  
 AC 09GK48  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Integrin alpha-V subunit (Fragment).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MAMMARY GLAND;  
 RA Andersen M.H., Rasmussen J.T., Berglund L., Petersen T.E.;  
 RT "Bovine alpha-V integrin subunit (fragment).";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AF317199; AAG38595.1; -  
 DR HSSP: P06756; IJV2.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR Pfam: PF01839; FG-GAP; 5.  
 DR Pfam: PF00357; Integrin\_A; 1.  
 DR PRINTS: PR01185; INTEGRINA.  
 DR SMART: SM00191; Int\_alpha; 5.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 FT NON TER 1  
 SQ SEQUENCE 1007 AA; 111460 MW; 51AFBC6DFCC36761 CRC64;

Query Match 43.9%; Score 50; DB 6; Length 1007;  
 Best Local Similarity 38.1%; Pred. No. 95;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFPAAKKIPEEKREKLE 21  
 Db 976 RMGFFKVRPQEGEREDQ 996

RESULT 11  
 09MZD6 PRELIMINARY; PRT; 1047 AA.  
 ID 09MZD6  
 AC 09MZD6  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Integrin alpha V subunit precursor.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RX MEDLINE=20366286; PubMed=10906183;  
 RA Neff S., Mason P.W., Baxt B.;  
 RT "High-efficiency utilization of the bovine integrin alpha(v)beta(3) as  
 RT a receptor for foot-and-mouth disease virus is dependent on the bovine  
 RT beta(3) subunit."  
 RL J. Virol. 74:7298-7306(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RA Neff S., Mason P.W., Baxt B.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF239958; AAF44691.2; -  
 DR HSSP: P06756; IJV2.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR Pfam: PF01839; FG-GAP; 5.  
 DR Pfam: PF00357; Integrin\_A; 1.  
 DR PRINTS: PR01185; INTEGRINA.  
 DR SMART: SM00191; Int\_alpha; 4.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 KW Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1047 INTEGRIN ALPHA V SUBUNIT.  
 SQ SEQUENCE 1047 AA; 116093 MW; 285F75F17471CA4B CRC64;

Query Match 43.9%; Score 50; DB 6; Length 1047;  
 Best Local Similarity 38.1%; Pred. No. 98;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFPAAKKIPEEKREKLE 21  
 Db 1016 RMGFFKVRPQEGEREDQ 1036

RESULT 12  
 064016 PRELIMINARY; PRT; 236 AA.  
 ID 064016  
 AC 064016  
 DT 01-AUG-1998 (Tremblrel. 07, Created)  
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE Putative lipoprotein.  
 GN YOKB.  
 OS Bacteriophage SPBC2.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 OX NCBI\_TaxID=66797;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Meuel C.,  
 RA Karamata D.;  
 RT "The complete nucleotide sequence of the Bacillus subtilis SPB2ac2  
 RT prophage.";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

```

DR EMBL AF020713; AAC12975.1; -.
KW Lipoprotein.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26CDCF4FAAC CRC64;

Query Match 43.4%; Score 49.5; DB 9; Length 236;
Best Local Similarity 84.6%; Pred. No. 28;
Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 8 KKI-PEEKREK 19
DB 205 KKVSPPEEKREK 217

RESULT 13
ID 032005 PRELIMINARY; PRT; 236 AA.
AC 032005;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE YOKB protein.
DE YOKB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunev F., Ogatawara N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Berteiro M.G., Beesler P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Britnell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabre C., Ferrati E., Foulger D.,
RA Fritz C., Fujita Y., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Goldlighty E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut L.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Jorle B., Karamata D., Kasahara Y., Kieffer-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kunano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Pariz V., Pohl T.M., Portetelle D., Potwilk S., Prescott A.M.,
RA Pirescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccetti E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandendri M., Vannier P., Vassarotti A.,
RA Viari A., Wambolt R., Wedler E., Wedler H., Weitzengesser T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yamumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunev F., Ogatawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z59115; CAB14083.1; -.
KW Complete proteome.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26CDCF4FAAC CRC64;

Query Match 43.4%; Score 49.5; DB 16; Length 236;
Best Local Similarity 84.6%; Pred. No. 28;
Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

```

```

QY 8 KKI-PEEKREK 19
DB 205 KKVSPPEEKREK 217

RESULT 14
ID 096LR8 PRELIMINARY; PRT; 635 AA.
AC 096LR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CDNA FLJ25142 f1s, clone CBR07252, highly similar to Homo sapiens
DE putative transcription factor-like nuclear regulator (TFNR gene)
DE (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Tashtiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Takizumi Y., Fujimori Y., Komiyama M., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF057871; BAB71602.1; -.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN.1.
FT NON_TER 635 635
SQ SEQUENCE 635 AA; 71552 MW; CC7302D845C941F2 CRC64;

Query Match 43.4%; Score 49.5; DB 4; Length 635;
Best Local Similarity 57.9%; Pred. No. 73;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 4 FFAH--KKIPEEKREK 19
DB 346 FFAHLLQKVLAEKREK 364

RESULT 15
ID 09HAW1 PRELIMINARY; PRT; 1372 AA.
AC 09HAW1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE RNA polymerase III transcription initiation factor B', short.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20496900; PubMed=11040218;
RA Schramm L., Pendergrast P.S., Sun Y., Hernandez N.;
RT "Different human TFIIIB activities direct RNA polymerase III
RT transcription from TATA-containing and TATA-less promoters."
RL Genes Dev. 14:2650-2663(2000).
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR EMBL; AF298152; AAG30221.1; -.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00395; SANT; 1.

```

DR PROSITE; PS00697; DNA LIGASE A1; UNKNOWN 1.  
KW DNA-binding; Initiation factor; Nuclear protein.  
SQ SEQUENCE 1372 AA; 154605 MW; C7416FD179610F22 CRC64;

Query Match 43.4%; Score 49.5; DB 4; Length 1372;  
Best Local Similarity 57.9%; Pred. No. 1.5e+02;  
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

OY 4 FFAH---KKIPEEKREK 19

DB 366 FFAHLLOKVLAEERKOK 384

Search completed: July 16, 2003, 07:55:30  
Job time : 4.25879 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:55:38 ; Search time 1.59353 Seconds  
(without alignments)  
1639.577 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114  
Sequence: 1 KLGFPAHKKIPPEEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	44.7	856	10	US-09-815-242-11310
2	49	43.0	856	10	US-10-083-357-1275
3	49	43.0	856	10	US-09-815-242-11489
4	46	40.4	479	10	US-09-971-309-64
5	44	38.6	19	9	US-10-101-487-67
6	44	38.6	20	9	US-09-520-907B-24
7	44	38.6	93	9	US-10-134-841-2
8	44	38.6	122	9	US-10-106-698-6119
9	44	38.6	176	9	US-10-101-487-70
10	44	38.6	179	9	US-10-101-487-46
11	44	38.6	186	9	US-10-101-487-44
12	44	38.6	187	9	US-10-101-487-50
13	44	38.6	191	9	US-10-101-487-81
14	44	38.6	198	9	US-10-101-487-42
15	44	38.6	200	9	US-09-736-457-789
16	44	38.6	200	9	US-10-101-487-53
17	44	38.6	200	9	US-09-902-941-788
18	44	38.6	200	9	US-09-849-626-789
19	44	38.6	200	9	US-10-017-754-789

20	44	38.6	207	9	US-09-736-457-1667	Sequence 1667, Ap
21	44	38.6	207	9	US-09-902-941-1667	Sequence 1667, Ap
22	44	38.6	207	9	US-09-902-941-1913	Sequence 1913, Ap
23	44	38.6	207	9	US-09-849-626-1667	Sequence 1667, Ap
24	44	38.6	207	9	US-09-849-626-1913	Sequence 1913, Ap
25	44	38.6	207	9	US-10-017-754-1667	Sequence 1667, Ap
26	44	38.6	207	9	US-10-017-754-1913	Sequence 1913, Ap
27	44	38.6	221	9	US-10-094-458A-9	Sequence 9, Ap
28	44	38.6	430	9	US-10-286-264-104	Sequence 104, App
29	44	38.6	430	9	US-10-094-458A-3	Sequence 3, App
30	44	38.6	430	9	US-10-094-458A-6	Sequence 6, App
31	44	38.6	430	9	US-10-295-403-34	Sequence 34, App
32	44	38.6	430	9	US-09-934-455-154	Sequence 154, App
33	44	38.6	461	9	US-10-234-432-24	Sequence 24, App
34	44	38.6	676	9	US-10-128-714-3166	Sequence 3166, Ap
35	44	38.6	676	9	US-10-128-714-8166	Sequence 8166, Ap
36	44	38.6	893	9	US-10-234-432-86	Sequence 86, App
37	43.5	38.2	1456	9	US-10-205-823-113	Sequence 113, App
38	43.5	38.2	1494	9	US-10-205-823-111	Sequence 111, App
39	43.5	38.2	1863	9	US-09-734-672-2	Sequence 2, App
40	43.5	38.2	1863	9	US-09-734-672-4	Sequence 4, App
41	43.5	38.2	1863	9	US-09-734-672-6	Sequence 6, App
42	43.5	38.2	1863	9	US-09-982-828-2	Sequence 2, App
43	43.5	38.2	1863	9	US-09-982-828-4	Sequence 4, App
44	43.5	38.2	1863	9	US-09-982-828-6	Sequence 6, App
45	43.5	38.2	1863	9	US-10-022-819-2	Sequence 2, App

## ALIGNMENTS

RESULT 1  
US-09-815-242-11310  
Sequence 11310, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11310  
LENGTH: 856  
TYPE: PRT  
ORGANISM: Helicobacter pylori  
US-09-815-242-11310  
Query Match 44.7%; Score 51; DB 10; Length 856;  
Best Local Similarity 45.0%; Pred. No. 61;

Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
QY 3 GFFAHKKPEEKREKLEQ 22  
DB 496 GEIYSKIPENKKKEBELQ 515

## RESULT 2

US-10-083-357-1275  
Sequence 1275, Application US/10083357  
Publication No. US20030054370A1  
GENERAL INFORMATION:  
APPLICANT: Qiandong Zeng et al.  
TITLE OF INVENTION: Systemic Discovery of New Genes  
FILE REFERENCE: 032796-090  
CURRENT FILING DATE: 2002-02-27  
NUMBER OF SEQ ID NOS: 1346  
SEQ ID NO 1275  
LENGTH: 413  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-083-357-1275

Query Match 43.0%; Score 49; DB 9; Length 413;  
Best Local Similarity 56.2%; Pred. No. 52;  
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 7 HKKPEEKREKLEQ 22  
DB 373 NKRIROKEREKLE 388

## RESULT 3

US-09-815-242-11489  
Sequence 11489, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT FILING DATE: 2001-03-21  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11489  
LENGTH: 856  
TYPE: PRT  
ORGANISM: Helicobacter pylori  
US-09-815-242-11489

Query Match 43.0%; Score 49; DB 10; Length 856;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
QY 7 HKKPEEKREKLE 21  
DB 500 YSKPEKEKKEKLEQ 514

## RESULT 4

US-09-971-309-64  
Sequence 64, Application US/09971309  
Patent No. US20020106675A1  
GENERAL INFORMATION:  
APPLICANT: UEMORI, Takashi  
APPLICANT: SATO, Yoshimi  
APPLICANT: FUJITA, Tomoko  
APPLICANT: MIYAKE, Kazuo  
APPLICANT: MUKAI, Hiroyuki  
APPLICANT: ASADA, Kiyozo  
APPLICANT: KATO, Ikunobu  
TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS  
FILE REFERENCE: 1422-0494P  
CURRENT FILING DATE: 2001-10-05  
CURRENT APPLICATION NUMBER: US 09/446,504  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: PCT/JP98/02845  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: JP 9-187496  
PRIOR FILING DATE: 1997-06-26  
PRIOR APPLICATION NUMBER: JP 9-320692  
PRIOR FILING DATE: 1997-11-27  
NUMBER OF SEQ ID NOS: 92  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 64  
LENGTH: 479  
TYPE: PRT  
ORGANISM: Pyrococcus furiosus  
US-09-971-309-64

Query Match 40.4%; Score 46; DB 10; Length 479;  
Best Local Similarity 75.0%; Pred. No. 1.6e+02;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 11 PEEKREKLEQ 22  
DB 451 PEEKREKLEK 462

## RESULT 5

US-10-101-487-67  
Sequence 67, Application US/10101487  
Patent No. US20020169125A1  
GENERAL INFORMATION:  
APPLICANT: LEUNG, DAVID W.  
APPLICANT: BERGMAN, PHILIP A.  
APPLICANT: LOBOUIST, ALAN  
APPLICANT: PIETZ, GREGORY E.  
APPLICANT: TOMPKINS, CHRISTOPHER K.  
APPLICANT: WAGGONER JR., DAVID W.  
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
FILE REFERENCE: 077319/0329  
CURRENT FILING DATE: 2002-03-20  
CURRENT APPLICATION NUMBER: US/10/101,487  
PRIOR FILING DATE: 2002-03-20  
PRIOR APPLICATION NUMBER: 60/277,705  
PRIOR FILING DATE: 2001-03-21  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 67  
LENGTH: 19  
TYPE: PRT

ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-10-101-487-67

Query Match 38.6%; Score 44; DB 9; Length 19;  
Best Local Similarity 57.1%; Pred. No. 9.5;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 KIPPEKREKLEQ 22  
|:|||||:|:|:  
Db 2 KMPPEEEEEEE 15

RESULT 6  
US-09-320-907B-24  
; Sequence 24, Application US/09320907B  
; Publication No. US20020198360A1  
; GENERAL INFORMATION:  
; APPLICANT: GINSBERG, MARK H.  
; APPLICANT: PFAFF, MARTIN  
; TITLE OF INVENTION: STRUCTURAL MODELS FOR CYTOPLASMIC DOMAINS OF  
; FILE REFERENCE: SRI-0006  
; CURRENT APPLICATION NUMBER: US/09/320,907B  
; CURRENT FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: 09/187,236  
; PRIOR FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-320-907B-24

Query Match 38.6%; Score 44; DB 9; Length 20;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPEEKKEE 18  
|:|||||:|:|:  
Db 1 KVGFFKRNRPPLLEAAEE 18

RESULT 7  
US-10-134-841-2  
; Sequence 2, Application US/10134841  
; Publication No. US20030003482A1  
; GENERAL INFORMATION:  
; APPLICANT: HALL, JOHN-PETER  
; APPLICANT: GOSPELT, ANDREAS  
; TITLE OF INVENTION: MRP8/MRP14 heterodimer, or its  
; TITLE OF INVENTION: individual components in combination, for treating and/or  
; TITLE OF INVENTION: preventing skin diseases, wounds and/or wound-healing  
; TITLE OF INVENTION: preventing skin diseases, having a reduced quantity of MRP8/MRP14  
; TITLE OF INVENTION: heterodimers  
; FILE REFERENCE: 50125/031002  
; CURRENT APPLICATION NUMBER: US/10/134,841  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: US 60/332,925  
; PRIOR FILING DATE: 2001-09-17  
; PRIOR APPLICATION NUMBER: DE 10121254.2  
; PRIOR FILING DATE: 2001-04-30  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 93  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-134-841-2

Query Match 38.6%; Score 44; DB 9; Length 93;  
Best Local Similarity 52.9%; Pred. No. 51;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPEEKKEE 17  
|:|||||:|:|:  
Db 77 KMGVAHKKSHESKKE 93

RESULT 8  
US-10-106-698-6119  
; Sequence 6119, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 6119  
; LENGTH: 122  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-106-698-6119

Query Match 38.6%; Score 44; DB 9; Length 122;  
Best Local Similarity 47.1%; Pred. No. 68;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREKLEQ 22  
|:|||||:|:|:  
Db 98 ARKKVEEDDEEEEEEE 114

RESULT 9  
US-10-101-487-70  
; Sequence 70, Application US/10101487  
; Patent No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOUIS, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/277,705  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 70  
; LENGTH: 176  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-10-101-487-70

Query Match 38.6%; Score 44; DB 9; Length 176;  
Best Local Similarity 57.1%; Pred. No. 1e+02;

Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
Qy 9 KIPEEKREKLEQ 22  
|:||||:|:|:  
Db 2 KMPEEEEEEEEE 15

RESULT 10  
US-10-101-487-46  
; Sequence 46, Application US/10101487  
; Patent No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOUJIST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/277,705  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 46  
; LENGTH: 179  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-10-101-487-46

Query Match 38.6%; Score 44; DB 9; Length 179;  
Best Local Similarity 57.1%; Pred. No. 1e+02;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 KIPEEKREKLEQ 22  
|:||||:|:|:  
Db 2 KMPEEEEEEEEE 15

RESULT 11  
US-10-101-487-44  
; Sequence 44, Application US/10101487  
; Patent No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOUJIST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/277,705  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 44  
; LENGTH: 186  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic fusion  
US-10-101-487-44

Query Match 38.6%; Score 44; DB 9; Length 186;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 KIPEEKREKLEQ 22  
|:||||:|:|:  
Db 2 KMPEEEEEEEEE 22

RESULT 12  
US-10-101-487-50  
; Sequence 50, Application US/10101487  
; Patent No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOUJIST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/277,705  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 50  
; LENGTH: 187  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-10-101-487-50

Query Match 38.6%; Score 44; DB 9; Length 187;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 KIPEEKREKLEQ 22  
|:||||:|:|:  
Db 2 KMPEEEEEEEEE 15

RESULT 13  
US-10-101-487-81  
; Sequence 81, Application US/10101487  
; Patent No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOUJIST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/277,705  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 81  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-10-101-487-81

Query Match 38.6%; Score 44; DB 9; Length 191;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 KIPPEEKREKLEQ 22  
|:||||:|:|:  
Db 2 KMPEEEEEEEEE 15

## RESULT 14

US-10-101-487-42  
; Sequence 42; Application US/10101487  
; Patent No. US20020169125A1  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, DAVID W.  
; APPLICANT: BERGMAN, PHILIP A.  
; APPLICANT: LOFOUISST, ALAN  
; APPLICANT: PIETZ, GREGORY E.  
; APPLICANT: TOMPKINS, CHRISTOPHER K.  
; APPLICANT: WAGGONER, JR., DAVID W.  
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
; FILE REFERENCE: 077319/0329  
; CURRENT APPLICATION NUMBER: US/10/101,487  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/277,705  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 42  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic fusion  
; OTHER INFORMATION: protein  
US-10-101-487-42

Query Match 38.6%; Score 44; DB 9; Length 198;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 KIPPEEKREKLEQ 22  
|:||||:|:|:  
Db 9 KMPEEEEEEEEE 22

## RESULT 15

US-09-736-457-789  
; Sequence 789; Application US/09736457  
; Patent No. US20020168637A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darrick  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Fan, Liqun  
; APPLICANT: Wang, Aijun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.478C15  
; CURRENT APPLICATION NUMBER: US/09/736,457  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 1864  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 789  
; LENGTH: 200  
; TYPE: PRT

; ORGANISM: Homo sapien  
US-09-736-457-789

Query Match 38.6%; Score 44; DB 9; Length 200;  
Best Local Similarity 47.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKIPPEEKREKLEQ 22  
|:||||:|:|:  
Db 176 ARKKVEEEDDEEEEEEE 192

Search completed: July 16, 2003, 08:12:53  
Job time : 2.59353 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:22 ; Search time 2.44444 Seconds  
(without alignments)  
1199.256 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114  
Sequence: 1 KLGFPAHKKIPBEKEKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: A.Geneseq\_101002: \*  
2: /SID2/gcgdata/geneseq/genesqp-emb1/AA1980.DAT: \*  
3: /SID2/gcgdata/geneseq/genesqp-emb1/AA1981.DAT: \*  
4: /SID2/gcgdata/geneseq/genesqp-emb1/AA1982.DAT: \*  
5: /SID2/gcgdata/geneseq/genesqp-emb1/AA1983.DAT: \*  
6: /SID2/gcgdata/geneseq/genesqp-emb1/AA1984.DAT: \*  
7: /SID2/gcgdata/geneseq/genesqp-emb1/AA1985.DAT: \*  
8: /SID2/gcgdata/geneseq/genesqp-emb1/AA1986.DAT: \*  
9: /SID2/gcgdata/geneseq/genesqp-emb1/AA1987.DAT: \*  
10: /SID2/gcgdata/geneseq/genesqp-emb1/AA1988.DAT: \*  
11: /SID2/gcgdata/geneseq/genesqp-emb1/AA1989.DAT: \*  
12: /SID2/gcgdata/geneseq/genesqp-emb1/AA1990.DAT: \*  
13: /SID2/gcgdata/geneseq/genesqp-emb1/AA1991.DAT: \*  
14: /SID2/gcgdata/geneseq/genesqp-emb1/AA1992.DAT: \*  
15: /SID2/gcgdata/geneseq/genesqp-emb1/AA1993.DAT: \*  
16: /SID2/gcgdata/geneseq/genesqp-emb1/AA1994.DAT: \*  
17: /SID2/gcgdata/geneseq/genesqp-emb1/AA1995.DAT: \*  
18: /SID2/gcgdata/geneseq/genesqp-emb1/AA1996.DAT: \*  
19: /SID2/gcgdata/geneseq/genesqp-emb1/AA1997.DAT: \*  
20: /SID2/gcgdata/geneseq/genesqp-emb1/AA1998.DAT: \*  
21: /SID2/gcgdata/geneseq/genesqp-emb1/AA1999.DAT: \*  
22: /SID2/gcgdata/geneseq/genesqp-emb1/AA2000.DAT: \*  
23: /SID2/gcgdata/geneseq/genesqp-emb1/AA2001.DAT: \*  
24: /SID2/gcgdata/geneseq/genesqp-emb1/AA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	22	AAV32244	Human integrin sub
2	114	100.0	1132	AAV32243	Human integrin sub
3	114	100.0	1167	AAV32242	Human integrin sub
4	114	100.0	1167	AAV32241	Human integrin sub
5	51	44.7	121	ABG26508	Novel secreted pro
6	51	44.7	147	ABG11269	Novel human diagno
7	51	44.7	502	ABG07742	Novel human diagno
8	51	44.7	856	AAU5717	Helicobacter pylori
9	50	43.9	32	AAU80955	Integrin alpha-V c
10	50	43.9	156	ABG18954	Novel human diagno

11	50	43.9	357	23	ABG48958
12	50	43.9	1048	23	AAU76335
13	50	43.9	1061	22	ABG18895
14	49	43.0	130	22	AAO00651
15	49	43.0	486	21	AAO3097
16	49	43.0	584	21	AAO3096
17	49	43.0	615	21	AAO3095
18	49	43.0	751	18	AAW13491
19	49	43.0	856	22	AAU5896
20	48.5	42.5	723	22	ABG28392
21	48	42.1	573	16	AAO8206
22	48	42.1	1464	23	ABG1199
23	47.5	41.7	2343	20	AAO80989
24	47.5	41.7	2343	20	AAO80989
25	47	41.2	136	22	ABG14329
26	47	41.2	177	21	AAO14466
27	47	41.2	177	21	AAO14466
28	47	41.2	193	22	ABG11252
29	47	41.2	211	21	AAO14465
30	47	41.2	211	21	AAO14465
31	47	41.2	306	21	AAO35906
32	47	41.2	702	22	AAO33203
33	47	41.2	702	22	AAO34499
34	47	41.2	1006	21	AAO33126
35	47	41.2	1409	21	AAO33125
36	47	41.2	1420	21	AAO33124
37	46	40.4	228	22	AAO6259
38	46	40.4	301	19	AAO80662
39	46	40.4	308	22	AAO1105
40	46	40.4	479	20	AAO7108
41	46	40.4	479	22	AAO62022
42	46	40.4	753	22	AAO30858
43	46	40.4	1615	22	AAO62021
44	46	40.4	2757	22	ABG09921
45	45	39.5	84	22	ABG11239

#### ALIGNMENTS

RESULT 1	
AAV32244	
ID	AAV32244 standard; Peptide; 22 AA.
AC	AAV32244;
DT	15-FEB-2000 (first entry)
XX	
XX	Human integrin subunit alpha-10 cytoplasmic domain peptide.
DE	
XX	Integrin alpha-10; ISa10; human; rheumatoid arthritis;
XX	osteoarthritis; osteoarthritis; cancer; atherosclerosis;
KW	inflammation; therapy; cartilage; chondrocyte; osteoblast;
KW	fibroblast; vaccine; marker.
XX	
OS	Homo sapiens.
XX	
PN	W09951639-A1;
PD	14-OCT-1999.
XX	
PF	31-MAR-1999; 99WO-SE00544.
XX	
PR	02-APR-1998; 98SE-0001164.
PR	28-JAN-1999; 99SE-0000319.
PA	(ACTI-) ACTIVE BIOTECH AB.
XX	
PI	Lundgren-Akerlund E;
XX	
DR	WPI; 2000-052639/04.
XX	
PT	New isolated integrin subunit alpha-10, used as a marker or target

PT molecule for cells during development, regeneration and pathological  
PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
PT inflammation

PS Claim 21; Page 53; 90pp; English.

XX This sequence represents a fragment of novel human chondrocyte  
CC integrin subunit alpha-10 (ISa10, see AAY32242), corresponding to  
CC the C-terminal cytoplasmic domain of the protein. The invention  
CC relates to a recombinant or isolated integrin heterodimer  
CC comprising the alpha10 subunit in association with subunit beta  
CC (especially beta-1). The integrin heterodimer, or the subunit  
CC alpha-10, or a fragment of it such as the present sequence, can be  
CC used as a marker or target of all types of cells, e.g. of  
CC chondrocytes, osteoblasts and fibroblasts. They can also be used:  
CC for treating pathological conditions involving ISa10, such as  
CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
CC for detecting the formation of cartilage during embryonal  
CC development, physiological or therapeutic repair of cartilage,  
CC or detecting regeneration of cartilage or chondrocytes during  
CC transplantation of cartilage or chondrocytes; for selection and  
CC analysis or for sorting, isolating or purification of chondrocytes  
CC and for in vitro studies of differentiation of chondrocytes; and as  
CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
CC skeletal muscle or other tissues where adhesion impairs the function  
CC of the tissue (all claimed). ISa10 binding entities can be used to  
CC determine the differentiation-state of cells during embryonic  
CC development, angiogenesis or development of cancer, in pathological  
CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,  
CC in tissue regeneration or in therapeutic and physiological repair  
CC of cartilage (claimed). A vaccine comprising the integrin  
CC heterodimer or subunit alpha-10 is also claimed. ISa10  
CC polynucleotides, vectors, host cells and methods of producing  
CC recombinant ISa10 are also claimed.

XX SQ Sequence 22 AA;

Query Match 100.0%; Score 114; DB 21; Length 22;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLGFFAHKKIPEEKREKLEQ 22  
|||  
1 KLGFFAHKKIPEEKREKLEQ 22

RESULT 2  
AAY32243 standard; Protein; 1132 AA.

XX AAY32243;

DT 15-FEB-2000 (first entry)

DE Human integrin subunit alpha-10 splice variant.

KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
KW inflammation; therapy; cartilage; chondrocyte; osteoblast;  
KW fibroblast; vaccine; marker; splice variant.

XX Homo sapiens.

OS Key Location/Qualifiers

XX Key

XX Peptide

XX Protein

FT 1..22 /note= "signal peptide"  
FT 23..1132 /note= "mature protein"

XX W0951639-A1.

XX 14-OCT-1999.

PF 31-MAR-1999; 99WO-SE00544.

XX 02-APR-1998; 98SE-0001164.

PR 28-JAN-1999; 98SE-0000319.

XX (ACTI-) ACTIVE BIOTECH AB.

PA Lundgren-Akerlund E;

XX WPI: 2000-052639/04.

DR N-PsDB: AA234720.

PT New isolated integrin subunit alpha-10, used as a marker or target  
PT molecule for cells during development, regeneration and pathological  
PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
PT inflammation

PS Claim 1; Page 43-48; 90pp; English.

XX This sequence represents a splice variant of novel human  
CC chondrocyte integrin subunit alpha-10 (ISa10). It is identical to  
CC ISa10 (see AAY32242) except for deletion of amino acids 975-986. The  
CC invention relates to a recombinant or isolated integrin heterodimer  
CC comprising the alpha10 subunit in association with subunit beta  
CC (especially beta-1). The heterodimer, subunit alpha-10 or splice  
CC variant can be used as a marker or target of all types of cells, e.g.  
CC of chondrocytes, osteoblasts and fibroblasts. They can also be used:  
CC for treating pathological conditions involving ISa10, such as  
CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
CC for detecting the formation of cartilage during embryonal  
CC development, physiological or therapeutic repair of cartilage,  
CC or detecting regeneration of cartilage or chondrocytes during  
CC transplantation of cartilage or chondrocytes; for selection and  
CC analysis or for sorting, isolating or purification of chondrocytes  
CC and for in vitro studies of differentiation of chondrocytes; and as  
CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
CC skeletal muscle or other tissues where adhesion impairs the function  
CC of the tissue (all claimed). ISa10 binding entities can be used to  
CC determine the differentiation-state of cells during embryonic  
CC development, angiogenesis or development of cancer, in pathological  
CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,  
CC in tissue regeneration or in therapeutic and physiological repair  
CC of cartilage (claimed). A vaccine comprising the integrin  
CC heterodimer or subunit alpha-10 is also claimed. ISa10  
CC polynucleotides, vectors, host cells and methods of producing  
CC recombinant ISa10 are also claimed.

XX SQ Sequence 1132 AA;

Query Match 100.0%; Score 114; DB 21; Length 1132;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLGFFAHKKIPEEKREKLEQ 22  
|||  
1111 KLGFFAHKKIPEEKREKLEQ 1132

RESULT 3  
AAY32242

ID AAY32242 standard; Protein; 1167 AA.

XX AAY32242;

DT 15-FEB-2000 (first entry)

DE Human integrin subunit alpha-10.

KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
KW inflammation; therapy; cartilage; chondrocyte; osteoblast;  
KW fibroblast; vaccine; marker.



OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FH Peptide 1..22  
 FT /note= "signal peptide"  
 FT 23..1145  
 FT /note= "mature protein"  
 FT 23..1120  
 FT Domain /note= "extracellular domain"  
 FT 1121..1145  
 FT /note= "transmembrane domain"  
 FT 1122..1167  
 FT /note= "cytoplasmic domain, specifically claimed  
 in Claim 21"  
 FT Domain 162..359  
 FT /note= "I-domain"  
 FT 494..502  
 FT Binding-site /note= "cation binding site motif"  
 FT 558..566  
 FT Binding-site /note= "cation binding site motif"  
 FT 620..628  
 FT Binding-site /note= "cation binding site motif"  
 FT 98  
 FT Modified-site /note= "N-glycosylated"  
 FT 336  
 FT Modified-site /note= "N-glycosylated"  
 FT 364  
 FT Modified-site /note= "N-glycosylated"  
 FT 733  
 FT Modified-site /note= "N-glycosylated"  
 FT 839  
 FT Modified-site /note= "N-glycosylated"  
 FT 921  
 FT Modified-site /note= "N-glycosylated"  
 FT 1018  
 FT Modified-site /note= "N-glycosylated"  
 FT 1039  
 FT Modified-site /note= "N-glycosylated"  
 FT /note= "N-glycosylated"  
 FT MO9951639-A1.  
 PN 14-OCT-1999.  
 PD 31-MAR-1999; 99WO-SE00544.  
 PF 02-APR-1998; 98SE-0001164.  
 PR 28-JAN-1999; 99SE-0000319.  
 PA (ACT1-) ACTIVE BIOTECH AB.  
 PI Lundgren-Akerlund E;  
 XX WPI: 2000-052639/04.  
 DR N-PSDB; AA234719.  
 XX  
 PT New isolated integrin subunit alpha-10, used as a marker or target  
 PT molecule for cells during development, regeneration and pathological  
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 PT inflammation -  
 PT  
 XX Claim 1; Fig 6; 90pp; English.  
 PS  
 XX This sequence represents novel human chondrocyte integrin subunit  
 CC alpha-10 (Isa10). A splice variant is given in AA32243. The  
 CC invention relates to a recombinant or isolated integrin heterodimer  
 CC comprising the alpha10 subunit in association with subunit beta  
 CC (especially beta-1). The heterodimer and the subunit alpha-10 can  
 CC be used as markers or targets of all types of cells, e.g. of  
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used:  
 CC for treating pathological conditions involving Isa10, such as  
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
 CC for detecting the formation of cartilage during embryonal  
 CC development, physiological or therapeutic reparation of cartilage,

CC or detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes; for selection and  
 CC analysis or for sorting, isolating or purification of chondrocytes  
 CC and for in vitro studies of differentiation of chondrocytes; and as  
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
 CC skeletal muscle or other tissues where adhesion impairs the function  
 CC of the tissue (all claimed). Isa10 binding entities can be used to  
 CC determine the differentiation-state of cells during embryonic  
 CC development, angiogenesis or development of cancer, in pathological  
 CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,  
 CC in tissue regeneration or in therapeutic and physiological reparation  
 CC of cartilage (claimed). A vaccine comprising the integrin  
 CC heterodimer or subunit alpha-10 is also claimed. Isa10  
 CC polynucleotides, vectors, host cells and methods of producing  
 CC recombinant Isa10 are also claimed.  
 CC  
 XX Sequence 1167 AA;  
 SQ  
 Query Match 100.0%; Score 114; DB 21; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KLGFPAHKKIPBEKREKLEQ 22  
 Db 1146 KLGFPAHKKIPBEKREKLEQ 1167  
 RESULT 4  
 ID AAB64584 standard; Protein; 1167 AA.  
 XX AAB64584;  
 AC 22-MAR-2001 (first entry)  
 XX  
 DT Human secreted protein #37.  
 XX  
 DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiatic; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200077197-A1.  
 PD 21-DEC-2000.  
 XX  
 PD 01-JUN-2000; 2000WO-US14934.  
 PF 11-JUN-1999; 99US-0138599.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis GA;  
 XX WPI: 2001-032312/04.  
 DR N-PSDB; AAF32793.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 PT  
 XX Claim 11; Page 496-500; 558pp; English.  
 PS  
 XX Sequences AAB64584-564594 represent the amino acid sequences of 47  
 CC human secreted proteins encoded by the genes AAF32757-F32803. The genes  
 CC and proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
 CC a range of human tissues disclosed in the specification. The nucleic  
 CC acids, proteins, antibodies and (ant)agonists are useful in the  
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and

CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.

XX  
 SQ Sequence 1167 AA;

Query Match 100.0%; Score 114; DB 22; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPPEEKREKLEQ 22  
 |||||  
 DB 1146 KLGFFAHKKIPPEEKREKLEQ 1167

RESULT 5  
 ABG26508  
 ID ABG26508 standard; Protein; 121 AA.  
 XX  
 AC ABG26508;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #26499.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Dmanac RT, Liu C, Tang YT;  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS90695.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID No 56867; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and

CC responsible for genetic disorders or other traits to assess biodiversity,  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 121 AA;

Query Match 44.7%; Score 51; DB 22; Length 121;  
 Best Local Similarity 55.6%; Pred. No. 10;  
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 FAHKKIPPEEKREKLEQ 22  
 |||||  
 DB 5 FLVKKIKKEEEEEEE 22

RESULT 6  
 ABG11269  
 ID ABG11269 standard; Protein; 147 AA.  
 XX  
 AC ABG11269;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #11260.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Dmanac RT, Liu C, Tang YT;  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS75456.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID No 41628; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and



CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

XX Sequence 856 AA;

Query Match 44.7%; Score 51; DB 22; Length 856;  
 Best Local Similarity 45.0%; Pred. No. 72;  
 Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 GFFAHKKIPEEKREKLEQ 22  
 Db 496 GEIEXSKIPEKKEELQR 515

# RESULT 9

AA80955  
 ID AAR80955 standard; peptide; 32 AA.

AC AAR80955;

XX 23-APR-1996 (first entry)

DE Integrin alpha-V cytoplasmic domain.

XX Integrin; chimeric; extracellular; transmembrane; cytoplasmic;  
 KW ligand; activation inhibitor; inflammation; autoimmune disease;  
 KW transplacental; thrombus; cancer.

XX Homo sapiens.

XX WO9525173-A1.

XX 21-SEP-1995.

XX 06-MAR-1995; 95WO-US02885.

XX 14-MAR-1994; 94US-0214770.

XX (SCRI) SCRIPPS RES INST.

XX Ginsberg MH, O'Toole TE;

XX WPI, 1995-336977/43.

XX Chimeric integrin mol. comprising reporter and target integrin  
 PT domains - for identifying integrin activation inhibitors which are  
 PT useful for treating or preventing unwanted immune responses

XX Disclosure; Figure 2; 50pp; English.

XX Chimeric integrin molecules comprising the extracellular and  
 CC transmembrane domains of a reporter integrin (RI) fused to the cyto-  
 CC plasmic domain of a target integrin (TI). The RI is pref. from  
 CC alpha1b-beta3 and the TI is pref. chosen from alphaV-beta3, alphaW-  
 CC beta2, alpha1-beta2, alpha2-beta1, alpha6-beta1, alpha6a-beta1,  
 CC alpha1b-beta3 or alpha4-beta1. This sequence is the cytoplasmic  
 CC domain of alpha-1b. The chimeric integrins are useful for  
 CC identifying TI inhibitors which can be used to treat mammalian  
 CC cancers, thrombosis or any unwanted immune response, e.g.  
 CC inflammation; autoimmune disease; allergies or organ/tissue  
 CC transplant rejection.

XX Sequence 32 AA;

Query Match 43.9%; Score 50; DB 16; Length 32;  
 Best Local Similarity 38.1%; Pred. No. 3.7;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 KLGFPAHKKIPEEKREKLE 21  
 Db 1 RMGFFKRVPRPQEQERERQ 21

RESULT 10  
 ABG18894  
 ID ABG18894 standard; Protein; 156 AA.

XX ABG18894;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #18885.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI, 2001-639362/73.

XX N-PSDB; AAS83081.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX Claim 20; SEQ ID No 49253; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 156 AA;

Query Match 43.9%; Score 50; DB 22; Length 156;  
 Best Local Similarity 38.1%; Pred. No. 18;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 KLGFPAHKKIPEEKREKLE 21  
 Db 125 RMGFFKRVPRPQEQERERQ 145

# RESULT 11

ABBA48958  
ID ABB48958 standard; Protein; 357 AA.  
XX  
XX ABB48958;  
AC  
XX  
XX 05-FEB-2002 (first entry)  
XX  
XX Listeria monocytogenes protein #1662.  
DE  
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.  
XX  
XX Listeria monocytogenes.  
OS  
XX WO200177335-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-FR01118.  
XX  
XX 11-APR-2000; 2000FR-0004629.  
XX  
XX (INSP) INST PASTEUR.  
XX  
XX Buchrieser C, Frangeul L, Couve E, Rusnick C, Fsihi H, Dehoux P,  
PI Dussurget O, Chetoui F, Nedjari H, Glaeer P, Kunst F, Coessart P,  
PI Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA,  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
PI Chakraborty T, Domann E, Hain T, Berche P, Chablit A, Durant L,  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
PI Madueno E, De Pablo B, Weiland J, Kaerst U, Entian K, Hauf J,  
PI Rose M, Voss H;  
XX  
XX WPI; 2002-010914/01.  
XX  
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides -  
XX  
XX  
XX Claim 6; SEQ ID No 1663; 192pp; French.  
XX  
XX The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABB403041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIFO  
CC at ftp.wifo.int/pub/published\_pct\_sequences.  
XX  
XX  
XX Sequence 357 AA;  
SQ  
Query Match 43.9%; Score 50; DB 23; Length 357;  
Best Local Similarity 71.4%; Pred. No. 42;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
OY 6 AHKXIPPEEKREK 19  
DB 146 AKKRPPEEKREK 159

AAU76335 standard; peptide; 1048 AA.  
ID AAU76335;  
XX  
XX  
XX 21-MAY-2002 (first entry)  
XX  
XX Human anti-dual integrin protein #1.  
XX  
XX Human; dual integrin; HC CDR; variable region;  
KW LC CDR; medical device; immune related disease; rheumatoid arthritis;  
KW gastric ulcer; asthma; allergic rhinitis; Crohn's pathology; arteriosclerosis;  
KW sickle cell anaemia; diabetes; cardiovascular disease; arteriosclerosis;  
KW atherosclerosis; restenosis; angina pectoris; myocardial infarction;  
KW infectious disease; pneumonia; leprosy; malaria; malignant disease;  
KW leukaemia; chronic myelocytic leukaemia; multiple myeloma;  
KW neurological disease; multiple sclerosis; Parkinson's disease;  
KW Alzheimer's disease; Creutzfeldt-Jakob disease.  
XX  
XX Homo sapiens.  
OS  
XX WO200212501-A2.  
XX  
XX 14-FEB-2002.  
XX  
XX 07-AUG-2001; 2001WO-US24784.  
XX  
XX 07-AUG-2000; 2000US-223363P.  
XX  
XX 01-AUG-2001; 2001US-0920267.  
XX  
XX (CENZ) CENTOCOR INC.  
XX  
XX Giles-Komar J, Heavner G, Snyder L, Trikha M;  
XX  
XX WPI; 2002-217193/27.  
XX  
XX Novel isolated mammalian anti-dual integrin antibody, useful for  
PT diagnosing or treating dual integrin related condition such as  
PT rheumatoid arthritis, gastric ulcer, asthma, arteriosclerosis,  
PT restenosis -  
XX  
XX  
XX Claim 81; Page 135-138; 144pp; English.  
XX  
XX The invention relates to an isolated mammalian anti-dual integrin  
CC antibody having at least one of the human heavy chain or light  
CC chain complementary determining region (CDR, HC CDR1-CDR3, LC  
CC CDR1-3). Also included are the nucleic acids encoding the CDRs,  
CC a vector comprising the nucleic acids, a host cell comprising the  
CC vector, an anti-idiotypic antibody that binds to the anti-dual  
CC integrin, a medical device comprising the antibody suitable for  
CC administration by parenteral, subcutaneous, intramuscular, intravenous,  
CC intracardiac, intrabronchial, intrabdominal, intracapsular,  
CC intracartilaginous, intracavitary, intracerebellar, or other  
CC routes as given in specification. The antibody is useful for diagnosing  
CC or treating a dual integrin related condition in an animal for example,  
CC immune related disease such as rheumatoid arthritis, gastric ulcer,  
CC asthma, allergic rhinitis, Crohn's pathology, sickle cell anaemia,  
CC diabetes, cardiovascular disease such as arteriosclerosis,  
CC atherosclerosis, restenosis, angina pectoris, myocardial infarction,  
CC infectious disease in a cell such as bacterial, viral, and fungal  
CC infections, pneumonia, leprosy, malaria, malignant disease such as  
CC leukaemia, chronic myelocytic leukaemia, Burkitt's lymphoma, multiple  
CC myeloma; neurological disease such as multiple sclerosis, Parkinson's  
CC disease, spinal ataxia, Alzheimer's disease, Creutzfeldt-Jakob  
CC disease and many other diseases given in the specification. The  
CC present sequence is an human dual integrin protein against which the  
CC antibodies of the invention were raised.  
XX  
XX  
XX Sequence 1048 AA;  
SQ  
Query Match 43.9%; Score 50; DB 23; Length 1048;  
Best Local Similarity 38.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;



```
XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 67568.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-012180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144633.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
```

```
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
Query Match 43.0%; Score 49; DB 21; Length 486;
Best Local Similarity 71.4%; Pred.No. 78;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 8 KKIPEEKKEKLE 21
Db 267 KKIPEEKKEKKE 280
```

Search completed: July 16, 2003, 07:52:08  
Job time : 3.44444 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 07:49:17 ; Search time 0.866385 Seconds  
(without alignments)  
747.132 Million cell updates/sec

Title: US-09-647-544-7  
Perfect score: 114  
Sequence: 1 KIGFFAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	43.9	32	1	US-08-214-770-2
2	50	43.9	32	5	PCT-US95-02885-2
3	48	42.1	573	2	US-08-745-934-4
4	48	42.1	573	4	US-09-147-009-7
5	47.5	41.7	2343	4	US-09-324-867-2
6	46	40.4	479	4	US-09-446-504-64
7	46	40.4	479	4	US-09-712-266-64
8	45	39.5	109	4	US-09-091-725-31
9	44	38.6	46	1	US-08-200-016-3
10	44	38.6	93	1	US-07-987-272A-7
11	44	38.6	93	1	US-07-987-272A-16
12	44	38.6	93	1	US-08-385-241-1
13	43.5	38.2	765	1	US-08-425-061-19
14	43.5	38.2	765	2	US-08-825-886-19
15	43.5	38.2	900	1	US-08-425-061-20
16	43.5	38.2	900	2	US-08-825-886-20
17	43.5	38.2	914	1	US-08-425-061-21
18	43.5	38.2	914	2	US-08-825-886-21
19	43.5	38.2	1202	1	US-08-425-061-22
20	43.5	38.2	1202	2	US-08-825-886-22
21	43.5	38.2	1363	1	US-08-425-061-23
22	43.5	38.2	1363	2	US-08-825-886-23
23	43.5	38.2	1852	1	US-08-425-061-24
24	43.5	38.2	1852	2	US-08-825-886-24
25	43.5	38.2	1863	1	US-08-425-061-16
26	43.5	38.2	1863	1	US-08-598-591-2
27	43.5	38.2	1863	1	US-08-480-784-2

28	43.5	38.2	1863	1	US-08-483-553-2	Sequence 2, Appli
29	43.5	38.2	1863	1	US-08-487-002-2	Sequence 2, Appli
30	43.5	38.2	1863	1	US-08-483-554B-2	Sequence 2, Appli
31	43.5	38.2	1863	1	US-08-798-691-2	Sequence 2, Appli
32	43.5	38.2	1863	1	US-08-798-691-4	Sequence 2, Appli
33	43.5	38.2	1863	1	US-08-798-691-6	Sequence 6, Appli
34	43.5	38.2	1863	1	US-08-488-011B-2	Sequence 2, Appli
35	43.5	38.2	1863	2	US-08-825-886-16	Sequence 16, Appli
36	43.5	38.2	1863	2	US-08-603-753D-2	Sequence 2, Appli
37	43.5	38.2	1863	3	US-08-825-487A-2	Sequence 2, Appli
38	43.5	38.2	1863	3	US-08-825-487A-4	Sequence 4, Appli
39	43.5	38.2	1863	3	US-08-825-487A-6	Sequence 6, Appli
40	43.5	38.2	1863	4	US-09-074-476-2	Sequence 2, Appli
41	43.5	38.2	1863	4	US-09-074-476-4	Sequence 4, Appli
42	43.5	38.2	1863	4	US-09-074-476-6	Sequence 6, Appli
43	43.5	38.2	1863	4	US-09-099-753-2	Sequence 2, Appli
44	43.5	38.2	1863	4	US-08-850-1727-2	Sequence 2, Appli
45	43.5	38.2	1863	4	US-08-986-106-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-214-770-2  
Sequence 2, Application US/08214770  
Patent No. 5523209  
GENERAL INFORMATION:  
APPLICANT: Ginsberg, Mark H.  
APPLICANT: O'Toole, Tim  
TITLE OF INVENTION: METHODS FOR IDENTIFYING  
TITLE OF INVENTION: INHIBITORS OF INTEGRIN  
TITLE OF INVENTION: ACTIVATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 50Z or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,770  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06410/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-214-770-2  
Query Match 43.9%, Score 50, DB 1, Length 32,  
Best Local Similarity 38.1%, Pred. No. 1.2,  
Matches 8, Conservative 7, Mismatches 0, Gaps 0.

QY 1 KLGFPAHKKIPPEEKREKLE 21  
:||||:||||:||||:  
Db 1 RMGFFKRVPRPOEGEREQLQ 21

## RESULT 2

PCT-US95-02885-2  
Sequence 2, Application PC/US9502885  
GENERAL INFORMATION:  
APPLICANT: Ginsberg, Mark H.  
APPLICANT: O'Toole, Timothy  
TITLE OF INVENTION: METHODS FOR IDENTIFYING INHIBITORS  
TITLE OF INVENTION: OF INTEGRIN ACTIVATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02885  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/214,770  
FILING DATE: March 14, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06410/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US95-02885-2

Query Match 43.9%; Score 50; DB 5; Length 32;  
Best Local Similarity 38.1%; Pred. No. 1.2;  
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 KLGFPAHKKIPPEEKREKLE 21  
:||||:||||:||||:  
Db 1 RMGFFKRVPRPOEGEREQLQ 21

## RESULT 3

US-08-745-934-4  
Sequence 4, Application US/08745934  
Patent No. 5861496  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: HUMAN SQUALENE EPOXIDASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA

ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/745,934  
FILING DATE: Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0151 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 855-0555  
TELEFAX: (415) 845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 573 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 1083804  
US-08-745-934-4

Query Match 42.1%; Score 48; DB 2; Length 573;  
Best Local Similarity 52.9%; Pred. No. 38;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 LGFPAHKKIPPEEKREK 18  
:||||:||||:||||:  
Db 75 IGFPAKSPPESEKKEQ 91

## RESULT 4

US-09-147-009-7  
Sequence 7, Application US/09147009  
Patent No. 6153815  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Process for Raising Squalene Levels in Plants  
NUMBER OF SEQUENCES: 11  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/147,009  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 573 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus norvegicus  
TISSUE TYPE: kidney  
CELL LINE: NRK  
IMMEDIATE SOURCE:  
LIBRARY: pcd2 library of H. Okayama  
CLONE: Tb-1

```

1      PUBLICATION INFORMATION:
2      AUTHORS: Sakakibara, J.
3      AUTHORS: Matanabe, R.
4      AUTHORS: Kanai, R.
5      AUTHORS: Ono, T.
6      TITLE: Molecular cloning and expression of rat
7      TITLE: sqalene epoxidase
8      JOURNAL: J. Biol. Chem.
9      VOLUME: 270
10     ISSUE: 1
11     PAGES: 17-20
12     DATE: 1995
13
14 US-09-147-009-7
15
16 Query Match      42.1%      Score 48; DB 4; Length 573;
17 Best Local Similarity 52.9%      Pred. No. 38;
18 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0.
19
20 QY      2 LGFFAHKKIPPEEKREE 18
21      :||| | |||:
22      75 IGFWAKSPSEKKEQ 91
23
24 RESULT 5
25 US-09-324-867-2
26 Sequence 2, Application US/09324867A
27 Patent No. 6251632
28 GENERAL INFORMATION:
29 APPLICANT: Lilligrap, David
30 APPLICANT: Cameron, Cherie
31 APPLICANT: No. 6251632ley, Colleen
32 APPLICANT: Horrocks, L. Suzanne Hoyle
33 APPLICANT: Hough, Christine
34 TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
35 FILE REFERENCE: 1669.0010002/JAG/BJD
36 CURRENT APPLICATION NUMBER: US/09/324,867A
37 CURRENT FILING DATE: 1999-06-03
38 EARLIER APPLICATION NUMBER: 09/035,141
39 EARLIER FILING DATE: 1998-03-059
40 EARLIER APPLICATION NUMBER: 60/039,953
41 EARLIER FILING DATE: 1997-03-06
42 NUMBER OF SEQ ID NOS: 63
43 SOFTWARE: PatentIn Ver. 2.0
44 SEQ ID NO 2
45 LENGTH: 2343
46 TYPE: PRT
47 ORGANISM: Canis familiaris
48 US-09-324-867-2
49
50 Query Match      41.7%      Score 47.5; DB 4; Length 2343;
51 Best Local Similarity 40.6%      Pred. No. 1.8e+02;
52 Matches 13; Conservative 3; Mismatches 3; Indels 13; Gaps 1.
53
54 OY      4 FFAP-----KKIPPEEKREEKLEQ 22
55      |||: |||||: |||:
56      1183 FFANLANYQENDTYNQEKKSPEEIRKRLTQ 1214
57
58 RESULT 6
59 US-09-446-504-64
60 Sequence 64, Application US/09446504
61 Patent No. 6218150
62 GENERAL INFORMATION:
63 APPLICANT: UEMORI, Takashi
64 APPLICANT: SATO, Yoshimi
65 APPLICANT: FUJITA, Tomoko
66 APPLICANT: MIYAKE, Kazue
67 APPLICANT: MUKAI, Hiroyuki
68 APPLICANT: ASADA, Kiyozo
69 APPLICANT: KATO, Ikunoshin
70 TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
71 FILE REFERENCE: 1422-408PCT
72 CURRENT APPLICATION NUMBER: US/09/446,504

```

```

CURRENT FILING DATE: 1999-12-23
PRIORITY APPLICATION NUMBER: PCT/JP98/02845
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: JP 9-187496
PRIOR FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: JP 9-320692
PRIOR FILING DATE: 1997-11-27
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 64
LENGTH: 479
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-09-446-504-64

Query Match          40.4%  Score 46;  DB 4;  Length 479;
Best Local Similarity 75.0%  Pred. NO. 60;
Matches          9;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0

QY      11  PEEKEEKKLEQ 22
      ||||| ||| :
Db      451 PEEKEEKKKEX 462

RESULT 7
US-09-712-266-64
Sequence 64, Application US/09712266
Patent No. 6333158
GENERAL INFORMATION:
APPLICANT: UEMORI, Takashi
APPLICANT: SATO, Yoshimi
APPLICANT: FUJITA, Tomoko
APPLICANT: MIYAKE, Kazuo
APPLICANT: MUKAI, Hiroyuki
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshi
TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
FILE REFERENCE: 1422-408PCT
CURRENT APPLICATION NUMBER: US/09/712,266
CURRENT FILING DATE: 2000-11-15
PRIOR APPLICATION NUMBER: US 09/446,504
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: PCT/JP98/02845
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: JP 9-187496
PRIOR FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: JP 9-320692
PRIOR FILING DATE: 1997-11-27
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 64
LENGTH: 479
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-09-712-266-64

Query Match          40.4%  Score 46;  DB 4;  Length 479;
Best Local Similarity 75.0%  Pred. NO. 60;
Matches          9;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0

QY      11  PEEKEEKKLEQ 22
      ||||| ||| :
Db      451 PEEKEEKKKEX 462

RESULT 8
US-09-091-725-31
Sequence 31, Application US/09091725
Patent No. 6329141
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Improved methods for transforming Phaffia
and recombinant DNA for use therein

```

NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster LLP  
STREET: 2000 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: United States of America  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/091,725  
FILING DATE: 23-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 95203620.0  
FILING DATE: 22-DEC-1995  
APPLICATION NUMBER: EP 96200943.7  
FILING DATE: 11-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: E. Victor Donahue  
REGISTRATION NUMBER: 35,492  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 109 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-091-725-31

Query Match 39.5%; Score 45; DB 4; Length 109;  
Best Local Similarity 52.6%; Pred. No. 19;  
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 3 GFFAHKIPPEEKRE 21  
DB 76 GASADASAPAEKKEKAE 94

RESULT 9  
US-08-200-016-3  
Sequence 3, Application US/08200016  
Patent No. 5614397  
GENERAL INFORMATION:  
APPLICANT: Weisman, Irving  
APPLICANT: Lagasse, Eric  
TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR MODULATING  
TITLE OF INVENTION: APOPTOSIS IN HEMATOLOGY CELLS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 2200 Sand Hill Road, Suite 100  
CITY: Menlo Park  
STATE: California  
COUNTRY: US  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/200,016  
FILING DATE: 22-FEB-1994  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Rae-Venter, Barbara  
REGISTRATION NUMBER: 32,750  
REFERENCE/DOCKET NUMBER: 06037/003001  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 854-5277  
TELEFAX: (415) 854-0875  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-200-016-3

Query Match 38.6%; Score 44; DB 1; Length 46;  
Best Local Similarity 52.9%; Pred. No. 11;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFPAHKIPPEEKRE 17  
DB 30 KMGVAHAKKSHESKHE 46

RESULT 10  
US-07-987-272A-7  
Sequence 7, Application US/07987272A  
Patent No. 5731166  
GENERAL INFORMATION:  
APPLICANT: Geczy, C., Simpson, R. J. and Lackmann, M  
TITLE OF INVENTION: No. 5731166el Chemotactic Factor  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cushman Darby & Cushman  
STREET: 1100 New York Avenue, N. W., Ninth Floor, East Tower  
CITY: Washington  
STATE: D. C.  
COUNTRY: USA  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/987,272A  
FILING DATE: 05-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PK 2127  
FILING DATE: 05-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PK 4463  
FILING DATE: 05-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Brinkman, David W  
REGISTRATION NUMBER: 20,817  
REFERENCE/DOCKET NUMBER: DMB/1925/200259  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861 3000  
TELEFAX: 202-822 0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-987-272A-7

Query Match 38.6%; Score 44; DB 1; Length 93;  
Best Local Similarity 52.9%; Pred. No. 22;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFPAHKIPPEEKRE 17  
DB 77 KMGVAHAKKSHESKHE 93

RESULT 11  
US-07-987-272A-16  
Sequence 16, Application US/07987272A  
Patent No. 5731166  
GENERAL INFORMATION:  
APPLICANT: Geczy, C., Simpson, R. J. and Lackmann, M  
TITLE OF INVENTION: No. 5731166el Chemotactic Factor  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cushman Darby & Cushman  
STREET: 1100 New York Avenue, N. W., Ninth Floor, East Tower  
CITY: Washington  
STATE: D. C.  
COUNTRY: USA  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/987,272A  
FILING DATE: 05-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PK 2127  
FILING DATE: 05-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PK 4463  
FILING DATE: 05-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Brinkman, David W  
REGISTRATION NUMBER: 20,817  
REFERENCE/DOCKET NUMBER: DMB/1925/200259  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861 3000  
TELEFAX: 202-822 0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-987-272A-16

Query Match 38.6%; Score 44; DB 1; Length 93;  
Best Local Similarity 52.9%; Pred. No. 22;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 KLGFPAAKKIPEEKRE 17  
Db 77 KMGVAHHKSHESHSKE 93

RESULT 12  
US-08-385-241-1  
Sequence 1, Application US/08385241  
Patent No. 5776348  
GENERAL INFORMATION:  
APPLICANT: Selengut Ph.D., Jeremy D.  
APPLICANT: Orme-Johnson Ph.D., William H.  
APPLICANT: Dreidler M.D., Stephen P.  
APPLICANT: Asakura M.D., Hirokaka  
TITLE OF INVENTION: SYSTEM AND METHOD FOR INHIBITING  
FORMATION OF CRYSTALLINE STRUCTURES THAT INCLUDE STRUVITE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Choate, Hall & Stewart  
STREET: 53 State Street  
CITY: Boston

STATE: MA  
COUNTRY: USA  
ZIP: 02109-2891  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/385,241  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Herschbach Ph.D., Brenda M.  
REGISTRATION NUMBER: P-39,223  
REFERENCE/DOCKET NUMBER: 492611-000 (MIT6915)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-5175  
TELEFAX: (617) 248-4000  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: hmrp-8 protein  
US-08-385-241-1

Query Match 38.6%; Score 44; DB 1; Length 93;  
Best Local Similarity 52.9%; Pred. No. 22;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 KLGFPAAKKIPEEKRE 17  
Db 77 KMGVAHHKSHESHSKE 93

RESULT 13  
US-08-425-061-19  
Sequence 19, Application US/08425061  
Patent No. 5622829  
GENERAL INFORMATION:  
APPLICANT: KING, Mary-Claire  
APPLICANT: FRIEDMAN, Lori  
APPLICANT: OSTERMEYER, Beth  
APPLICANT: ROWELL, Sarah  
APPLICANT: LYNCH, Eric  
APPLICANT: SZABO, Csilla  
APPLICANT: LEE, Ming  
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
CANCER  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,061  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 765 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-425-061-19

Query Match 38.2%; Score 43.5; DB 1; Length 765;  
Best Local Similarity 58.8%; Pred. No. 2.1e+02;  
Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 5 FAHKIPPEEKREKLE 21  
| : : | | | | | | | |  
DB 721 FVNPSLPREK-EKLE 736

RESULT 14  
US-08-825-886-19  
Sequence 19, Application US/08825886  
Patent No. 5821328  
GENERAL INFORMATION:  
APPLICANT: KING, Mary-Claire  
APPLICANT: FRIEDMAN, Lori  
APPLICANT: OSTERMEYER, Beth  
APPLICANT: ROWELL, Sarah  
APPLICANT: LYNCH, Eric  
APPLICANT: SZABO, Csilla  
APPLICANT: LEE, Ming  
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
TITLE OF INVENTION: CANCER  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/825,886  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/425,061  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 765 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-825-886-19

Query Match 38.2%; Score 43.5; DB 2; Length 765;  
Best Local Similarity 58.8%; Pred. No. 2.1e+02;  
Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 5 FAHKIPPEEKREKLE 21  
| : : | | | | | | | |  
DB 721 FVNPSLPREK-EKLE 736

RESULT 15  
US-08-425-061-20  
Sequence 20, Application US/08425061  
Patent No. 5622829  
GENERAL INFORMATION:  
APPLICANT: KING, Mary-Claire  
APPLICANT: FRIEDMAN, Lori  
APPLICANT: OSTERMEYER, Beth  
APPLICANT: ROWELL, Sarah  
APPLICANT: LYNCH, Eric  
APPLICANT: SZABO, Csilla  
APPLICANT: LEE, Ming  
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN  
TITLE OF INVENTION: CANCER  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,061  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 900 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-425-061-20

Query Match 38.2%; Score 43.5; DB 1; Length 900;  
Best Local Similarity 58.8%; Pred. No. 2.4e+02;  
Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 5 FAHKIPPEEKREKLE 21  
| : : | | | | | | | |  
DB 721 FVNPSLPREK-EKLE 736

Search completed: July 16, 2003, 07:57:49  
Job time: 1.86639 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:11:08 ; Search time 15 Seconds  
(without alignments)  
140.997 Million cell updates/sec

Title: US-09-647-544-7  
Perfect score: 114  
Sequence: 1 KLGFPAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 4263

Minimum DB seq length: 0  
Maximum DB seq length: 22

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	35.1	19	2	A49725 valine-tRNA ligase
2	30	26.3	20	2	neurofilament heav
3	29	25.4	17	2	enamelin, 26K - bo
4	29	25.4	20	2	cytochrome-c oxida
5	28	24.6	10	2	S65385 cytochrome-c oxida
6	27	23.7	8	2	I64832 Ca2+-transporting
7	27	23.7	15	2	PA0041 plastocyanin-plast
8	27	23.7	16	2	S13898 alkaline phosphata
9	27	23.7	17	2	G83975 hypotrichal prote
10	27	23.7	18	2	A61220 epsilon receptor m
11	27	23.7	20	2	S11416 ribosomal protein
12	26	22.8	15	2	S36891 ribosomal protein
13	26	22.8	18	2	E33178 78K heat shock pro
14	26	22.8	21	2	PC7043 ubiquitin carboxyl
15	26	22.8	22	2	P33084 ribosomal protein
16	26	22.8	22	2	PC4388 sulfite reductase
17	25	21.9	16	2	A35552 caldesmon - turkey
18	25	21.9	16	2	JH0517 insulin-like growt
19	25	21.9	18	2	S39845 2-aminobenzoate-Co
20	25	21.9	20	2	D49164 chromogranin-B - r
21	25	21.9	20	2	S06149 photosystem I chai
22	25	21.9	20	2	S77983 cytochrome-c oxida
23	25	21.9	21	2	SX0078 alanine dehydrogen
24	24.5	21.5	22	2	S00189 molitin - dog (ten
25	24	21.1	14	2	S65392 cytochrome-c oxida
26	24	21.1	18	2	B44995 alkanal monooxygen
27	24	21.1	19	2	A37968 neutral surface pro
28	24	21.1	22	2	H86433 protein T17H7.9 (l
29	23.5	20.6	19	2	S02808 nucleolin - bovine

30	23	20.2	14	2	A49018 myosin heavy chain
31	23	20.2	15	2	C37765 hypotrichal prote
32	23	20.2	18	2	I40062 shikimate 5-dehydr
33	23	20.2	19	2	S69166 ferredoxin b - Jap
34	23	20.2	20	2	JP0070 ribosomal protein
35	23	20.2	20	2	S06150 photosystem I chai
36	23	20.2	20	2	S19618 globin - polychaet
37	23	20.2	20	2	A15156 lectin, galactose/
38	23	20.2	21	2	S78416 ribosomal protein
39	23	20.2	22	2	JP0071 ribosomal protein
40	23	20.2	22	2	PC4330 ribosomal protein
41	23	20.2	22	2	S48136 lucosyltransferase
42	23	20.2	22	2	S78007 fucosyltransferase
43	22.5	19.7	20	2	A05310 apolipoprotein B -
44	22	19.3	10	2	S43625 cytochrome-c oxida
45	22	19.3	12	2	S16335 beta-conglycinin a

## ALIGNMENTS

## RESULT 1

A49725  
valine-tRNA ligase (BC 6.1.1.9) - rabbit (fragment)  
N.Alternate names: valyl-tRNA synthetase  
C.Species: Oryctolagus cuniculus (domestic rabbit)  
C.Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 03-Jun-2002  
A.Accession: A49725  
R.Bec, G., Kerjan, P., Waller, J.P.  
J. Biol. Chem. 269, 2086-2092, 1994  
A.Title: Reconstitution in Vitro of the valyl-tRNA synthetase-elongation factor (EF) 1be  
A.Reference number: A49725; MUID:94124563; PMID:8294461  
A.Accession: A49725  
A.Status: preliminary  
A.Molecule type: protein  
A.Residues: 1-19 <BEC>  
C.Keywords: aminocyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 35.1%; Score 40; DB 2; Length 19;

Best local similarity 50.0%; Pred. No. 26;

Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 9 KIPEEKREKLEQ 22  
DB 2 QLPXKAKNEKLEK 15

## RESULT 2

.153671  
neurofilament heavy subunit - human (fragment)  
C.Species: Homo sapiens (man)  
C.Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
A.Accession: 153671  
R.Figlewicz, D.A.; Rouleau, G.A.; Krizus, A.; Jullien, J.P.  
Gene 132, 297-300, 1993  
A.Title: Polymorphism in the multi-phosphorylation domain of the human neurofilament hea  
A.Reference number: 153671; MUID:94040777; PMID:8224877  
A.Accession: 153671  
A.Status: preliminary  
A.Molecule type: mRNA  
A.Residues: 1-20 <RES>  
A.Cross-references: GB:S66408; NID:G452861; PIDV:AA826609.1; PID:G452862  
C.Genetics:  
A.Gene: GDB:NEFH  
A.Cross-references: GDB:120225; OMIM:162230  
A.Map position: 22q12.1-22q13.1

Query Match 26.3%; Score 30; DB 2; Length 20;

Best local similarity 53.8%; Pred. No. 6.5e+02;

Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 9 KIPEEKREKLEQ 21

Db 6 KSPKAKSPKKE 18

## RESULT 3

S10786  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 19-Mar-1997 #sequence\_revision 21-Nov-1998 #text\_change 21-Nov-1998  
C:Accession: S10786  
R:Strawich, E.; Glimcher, M.J.  
E: J. Biochem. 191, 47-56, 1990  
A>Title: Tooth 'enamelins' identified mainly as serum proteins. Major 'enamelin' is albu  
A:Reference number: S10780; PMID:90336641; PMID:2379503  
A:Accession: S10786  
A:Molecule type: protein  
A:Residues: 1-17 <STR>  
C:Keywords: enamel; phosphoprotein

Query Match 25.4%; Score 29; DB 2; Length 17;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 HKKIP 11  
Db 10 HKKIP 14

## RESULT 4

S43627  
cytochrome-c oxidase (EC 1.9.3.1) chain Vb-H - trout (fragment)  
C:Species: Salmo sp. (trout)  
C>Date: 19-Mar-1997 #sequence\_revision 01-Aug-1997 #text\_change 02-Jul-1998  
C:Accession: S43627  
R:Freund, R.; Kadenbach, B.  
E: J. Biochem. 221, 1111-1116, 1994  
A>Title: Identification of tissue-specific isoforms for subunits Vb and VIIa of cytochr  
A:Reference number: S43624; PMID:94237550; PMID:8181469  
C:Accession: S43627  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-20 <PRE>  
C:Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory

Query Match 25.4%; Score 29; DB 2; Length 20;  
Best Local Similarity 40.0%; Pred. No. 8.9e+02;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 8 KKIPEEKREKLEQ 22  
Db 4 KGIPTDEGATGLEE 18

## RESULT 5

S65385  
cytochrome-c oxidase (EC 1.9.3.1) chain VIIa, hepatic - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 07-May-1999  
C:Accession: S65385  
R:Schaeffer, H.; Noack, H.; Halangsk, W.; Brandt, U.; von Jagow, G.  
E: J. Biochem. 230, 235-241, 1995  
A>Title: Cytochrome-c oxidase in developing rat heart. Enzymic properties and amino-term  
A:Reference number: S65372; PMID:95524529; PMID:7601105  
A:Accession: S65385  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-10 <SCH>  
C:Keywords: oxidoreductase

Query Match 24.6%; Score 28; DB 2; Length 10;  
Best Local Similarity 57.1%; Pred. No. 6.1e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 9 KIPEEK 15  
Db 4 KVEPKOK 10

## RESULT 6

I64832  
Ca2+-transporting ATPase (EC 3.6.3.8) - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 19-Apr-2002  
C:Accession: I64832  
R:Wu, K.  
A: J. Physiol. 264, 333-341, 1993  
A>Title: Molecular cloning and quantification of mRNA and protein encoding Ca2+-ATPase 1  
A:Reference number: I51892  
A:Accession: I64832  
A>Status: preliminary; translated from GB/EMBL/DBU  
A:Molecule type: mRNA  
A:Residues: 1-8 <RES>  
A:Cross-references: GB:M99223; NID:G203644; PIDN:AAA40992.1; PID:G203646  
C:Genetics:  
A:Gene: SERCALB  
C:Keywords: hydrolase

Query Match 23.7%; Score 27; DB 2; Length 8;  
Best Local Similarity 57.1%; Pred. No. 2.8e+05;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 11 PEEKRE 17  
Db 2 PEDERRK 8

## RESULT 7

PA0041  
plastocyanin-plastocyanin reductase (EC 1.10.99.1) - Arabidopsis thaliana (fragment)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 30-Jun-1992 #sequence\_revision 06-Jan-1995 #text\_change 03-Jun-2002  
C:Accession: PA0041  
R:Kamo, M.; Kawakami, T.; Miyatake, N.; Tsugita, A.  
A:Submitted to JIPID, July 1994  
A:Description: Separation and characterization of Arabidopsis proteins by two-dimensiona  
A:Reference number: PA0001  
A:Accession: PA0041  
A:Molecule type: protein  
A:Residues: 1-15 <KAM>  
A:Experimental source: leaf  
C:Keywords: oxidoreductase

Query Match 23.7%; Score 27; DB 2; Length 15;  
Best Local Similarity 50.0%; Pred. No. 1.3e+03;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 9 KIPEEK 16  
Db 8 RVPDMERK 15

## RESULT 8

S13898  
alkaline phosphatase (EC 3.1.3.1) - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 01-Aug-1997  
C:Accession: S13898  
R:Fujimori-Arai, Y.; Koyama, T.; Hirano, K.; Sakagishi, Y.; Komoda, T.  
A: J. Biochem. Biophys. 284, 330-325, 1991  
A>Title: Purification and partial characterization of intestinal-like alkaline phosphata  
A:Reference number: S13898; PMID:9112827; PMID:1989515  
A:Accession: S13898  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-16 <FTU>  
C:Keywords: phosphoric monoester hydrolase



Query Match 23.7%; Score 27; DB 2; Length 16;  
Best Local Similarity 83.3%; Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 10 IPEEEK 15  
DB 2 IPEEEK 7

## RESULT 9

G83975  
hypothetical protein BH2607 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: G83975

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

C:Accession: G83975

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-17 <STO>

A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA806326.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2607

Query Match 23.7%; Score 27; DB 2; Length 17;  
Best Local Similarity 50.0%; Pred. No. 1.4e+03;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 8 KRIPEEKREKLE 21  
DB 2 KEIPTKMKRIKDE 15

## RESULT 10

A61220

epsilon receptor modulating protein (EC 3.4.21.-) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Mar-1996

C:Accession: A61220

R:Matsumita, S.; Katz, D.H.

Cell. Immunol. 137, 252-259, 1991

A:Title: The murine epsilon receptor modulating protein: a novel serine protease which m

A:Reference number: A61220; MUID:91356570; PMID:1679381

C:Accession: A61220

A:Molecule type: protein

A:Residues: 1-18 <MAT>

C:Comment: This serine proteinase from a T cell hybridoma does not reduce levels of CD23

C:Keywords: hydrolase; serine proteinase

Query Match 23.7%; Score 27; DB 2; Length 18;  
Best Local Similarity 33.3%; Pred. No. 1.5e+03;  
Matches 4; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 8 KRIPEEKREK 19  
DB 2 KPAPKKEKKK 13

## RESULT 11

S11416

ribosomal protein L6, cytosolic [validated] - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 21-Jul-2000

C:Accession: S11416

R:Wittmann-Liebold, B.; Geiseler, A.W.; Lin, A.; Wool, I.G.

J. Supramol. Struct. 12, 425-433, 1979

A:Title: Sequence of the amino-terminal region of rat liver ribosomal proteins S4, S6, S

A:Reference number: S11413; MUID:80252792; PMID:398910

A:Accession: S11416  
A:Molecule type: protein  
A:Residues: 1-20 <MIT>  
A:Note: the protein is designated as ribosomal protein L6  
C:Keywords: protein biosynthesis; ribosome

Query Match 23.7%; Score 27; DB 2; Length 20;  
Best Local Similarity 35.7%; Pred. No. 1.7e+03;  
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKRIPEEKREK 19  
DB 1 AGEKAEKPKKEOK 14

## RESULT 12

S36891  
ribosomal protein - Mycobacterium bovis (fragment)

C:Species: Mycobacterium bovis

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Jan-1995

C:Accession: S36891

R:Ohara, N.; Kimura, M.; Higashi, Y.; Yamada, T.

FEBS Lett. 331, 9-14, 1993

A:Title: Isolation and amino acid sequence of the 30S ribosomal protein S19 from Mycobac

A:Reference number: S36887; MUID:94009653; PMID:8405418

C:Accession: S36891

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-15 <OHA>

Query Match 22.8%; Score 26; DB 2; Length 15;  
Best Local Similarity 33.3%; Pred. No. 1.7e+03;  
Matches 4; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 8 KRIPEEKREK 19  
DB 3 KNVPANSRRKAK 14

## RESULT 13

E33178

76k heat shock protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 31-May-1991 #sequence\_revision 31-May-1991 #text\_change 23-Mar-1993

C:Accession: E33178

R:Ward, L.D.; Hong, J.; Whitehead, R.H.; Simpson, R.J.

Electrophoresis 11, 883-891, 1990

A:Title: Development of a database of amino acid sequences for human colon carcinoma pro

A:Reference number: A33178; MUID:91176935; PMID:2079031

C:Accession: E33178

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-18 <MAR>

Query Match 22.8%; Score 26; DB 2; Length 18;  
Best Local Similarity 57.1%; Pred. No. 2.1e+03;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 12 EEEKREE 18  
DB 2 EEDKKED 8

## RESULT 14

PC7043

ubiquitin carboxyl-terminal hydrolase (EC 3.1.1.-) homolog - yeast (Pichia anomala) (fra

N:Alternate names: ALX1 protein

C:Species: Pichia anomala, Candida pelliculosa

C:Date: 04-Mar-2000 #sequence\_revision 04-Mar-2000 #text\_change 11-May-2000

C:Accession: PC7043

R:Sakajo, S.; Minagawa, N.; Yoshimoto, A.

Biosci. Biotechnol. Biochem. 63, 1889-1894, 1999

A:Title: Structure and regulatory expression of a single copy alternative oxidase gene f

A:Reference number: JC7145; MUID:20101471; PMID:10655554  
A:Accession: PC7043  
A:Molecule type: DNA  
A:Residues: 1-21 <SAs>  
A:Cross-references: DDBJ:AB026726  
C:Genetics:  
A:Gene: alx1  
C:Keywords: hydrolase

Query Match	22.8%	Score 26	DB 2	Length-21
Best Local Similarity	50.0%	Pred. No. 2.4e+03		
Matches	5	Conservative	3	Mismatches 2
				Indels 0
				Gaps 0

Qy	13	EKREKLEQ	22
		: : :	:
Db	12	EDKREGEKEE	21

RESULT 15

ribosomal protein L18 - Haloferax mediterranei (fragment)  
N/Alternate names: ribosomal protein HL13  
C/Species: Haloferax mediterranei  
C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 18-Jun-1993  
C/Accession: F33084  
R/McDougal, J.  
submitted to the Protein Sequence Database, June 1990  
A/Reference number: A33084  
A/Accession: F33084  
A/Molecule type: protein  
A/Residues: 1-22 <MCD>  
A/Experimental source: strain DSM 1411  
C/Superfamily: rat ribosomal protein L5  
C/Keywords: protein biosynthesis; ribosome

Query Match	22.8%;	Score 26;	DB 2;	Length 22;
Best Local Similarity	30.8%;	Pred. No. 2.5e+03;		
Matches	4;	Conservative	4;	Mismatches 5;
				Indels .0;
				Gaps 0;

```
QY      9 KIPPEEKREEKLE 21
         |::|::|:
Db      7 KVPMMRRRREVRTD 19
```

Search completed: July 16, 2003, 08:14:42  
Job time : 17 secs

NCBI TaxID=8730;  
Viperidae; Crotalini;  
OX

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

```

RN [1]
RP SEQUENCE.
RC TISSUE-Venom;
RX MEDLINE=97238898; PubMed=9083096;
RA Torii S., Naito M., Tsuno T.;
RT "Apoxin I, a novel apoptosis-inducing factor with L-amino acid oxidase
RL activity purified from Western diamondback rattlesnake venom.";
RJ J. Biol. Chem. 272:9539-9542(1997).
CC -1- FUNCTION: CATALYZES AN OXIDATIVE DEAMINATION OF PREDOMINANTLY
CC HYDROPHOBIC AND AROMATIC L-AMINO ACIDS. HAS AN ANTIBACTERIAL
CC EFFECT AND AN ABILITY TO INDUCE APOPTOSIS. THE H(2)O(2) PRODUCED
CC BY L-AMINO ACID OXIDATION IS INVOLVED IN THE APOXIN-I INDUCED
CC APOPTOSIS AND HEMORRHAGE CAUSED BY THE VENOM.
CC -1- CATALYTIC ACTIVITY: An L-amino acid + H(2)O + O(2) = a 2-oxo acid
CC + NH(3) + H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: PLASMA MEMBRANE (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE FLAVIN MONOAMINE OXIDASE FAMILY.
CC STRONG, TO MOUSE FIG-1.
KM Oxidoreductase: Flavoprotein; FAD; Venom; Apoptosis; Antibiotic.
FT NON TER 20
SQ SEQUENCE 20 AA; 2521 MW; 95E973C6C86588E7 CRC64;

Query Match 23.2%; Score 26.5; DB 1; Length 20;
Best Local Similarity 44.4%; Pred. No. 8e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 6 AHKIP-EKKREKLEQ 22
DB 1 AHDRLPLEXPRETDYEE 18

RESULT 3
GR78_HORSE STANDARD; PRT; 15 AA.
AC P16392;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 78 kDa glucose-regulated protein (GRP 78) (Immunoglobulin heavy chain
DE binding protein) (BiP) (Fragment).
GN HSPA5 OR GRP78.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE.
RX MEDLINE=90147817; PubMed=1689156;
RA Obias B., Boyd N.D., Lubber-Narod J., Reyes V.E., Leeman S.E.;
RT "Isolation and identification of a polypeptide in the Hsp 70 family
RT that binds substance P.";
RL Biochem. Biophys. Res. Commun. 166:978-983(1990).
CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF
CC MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR InterPro: IPR001023; Hsp70.
DR PROSITE: PS00297; HSP70_1; PARTIAL.
DR PROSITE: PS00329; HSP70_2; PARTIAL.
DR PROSITE: PS01036; HSP70_3; PARTIAL.
KM ATP-binding; Endoplasmic reticulum.
FT NON TER 15
SQ SEQUENCE 15 AA; 1647 MW; 81119D21D0EC26DB CRC64;

Query Match 22.8%; Score 26; DB 1; Length 15;
Best Local Similarity 57.1%; Pred. No. 7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 12 EKKREE 18
DB 2 EKKED 8

```

```

RESULT 4
RL18_HALME STANDARD; PRT; 22 AA.
ID RL18_HALME
AC P50561;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L18P (HmeL18) (Fragment).
GN RPL18P.
OS Halobacterium mediterranei (Haloferax mediterranei).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloferax.
OX NCBI_TaxID=2252;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 1411;
RX MEDLINE=94229075; PubMed=8174557;
RA McDougall J., Wilmann-Liebold B.;
RT "Comparative analysis of the protein components from 5S rRNA-protein
RT complexes of halophilic archaeobacteria.";
RL Eur. J. Biochem. 221:779-785(1994).
CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
CC PIR: F33084; F33084.
KM Ribosomal protein.
FT NON TER 22
SQ SEQUENCE 22 AA; 2773 MW; FE1F607FCAE9876D CRC64;

Query Match 22.8%; Score 26; DB 1; Length 22;
Best Local Similarity 30.8%; Pred. No. 1e+03;
Matches 4; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 9 KIPPEKREKLE 21
DB 7 KVPMPRRREVRTD 19

RESULT 5
MCA2_RHOOP STANDARD; PRT; 15 AA.
ID MCA2_RHOOP
AC P56870;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative malylacetate reductase II (EC 1.3.1.32) (Fragment).
OS Rhodococcus opacus (Nocardia opaca).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=37919;
RN [1]
RP SEQUENCE.
RC STRAIN=IC;
RX MEDLINE=98324954; PubMed=9657989;
RA Seibert V., Koubatova E.M., Golovleva L.A., Schloemann M.;
RT "Characterization of a malylacetate reductase encoding region from
RT Rhodococcus opacus 1CP.";
RL J. Bacteriol. 180:3503-3508(1998).
CC -1- CATALYTIC ACTIVITY: 3-oxoadipate + NAD(P)(+) = 2-malylacetate +
CC NAD(P)H.
CC -1- PATHWAY: 3-CHLOROCATECHOL DEGRADATION (BETA-KETOADIPATE PATHWAY).
CC THIS PATHWAY SERVES A VITAL ROLE IN THE BIODEGRADATION OF TOXIC
CC AROMATIC COMPOUNDS INTRODUCED IN THE ENVIRONMENT BOTH AS NATURAL
CC PRODUCTS AND AS INDUSTRIAL EFFLUENT.
CC -1- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR InterPro: IPR001670; Fe-ADH.
DR PROSITE: PS00913; ADH_IRON_1; PARTIAL.
DR PROSITE: PS00060; ADH_IRON_2; PARTIAL.
KM Aromatic hydrocarbons catabolism; Oxidoreductase; NAD.
FT NON TER 15
SQ SEQUENCE 15 AA; 1884 MW; 58DA9DD038F025E CRC64;

```

Query Match 21.9%; Score 25; DB 1; Length 15;  
 Best Local Similarity 37.5%; Pred. No. 9.6e+02;  
 Matches 3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 FAHKIPEE 12  
 DB 3 FEHENDPO 10

## RESULT 6

ID 1BP4\_PIG STANDARD; PRT; 16 AA.  
 AC P24654;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Insulin-like growth factor binding protein 4 (IGFBP-4) (1BP-4)  
 DE (IGF-binding protein 4) (Fragment).  
 GN IGFBP4.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=96823;  
 RN [1]

RP SEQUENCE.  
 RX MEDLINE=92109718; PubMed=1722398;  
 RA Coleman M.E., Pan Y.-C.E., Ethernon T.D.;  
 RT "Identification and NH<sub>2</sub>-terminal amino acid sequence of three  
 insulin-like growth factor-binding proteins in porcine serum.";  
 RL Biochem. Biophys. Res. Commun. 181:1131-1136(1991).  
 CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs  
 AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
 PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
 INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC PROTEIN FAMILY.  
 CC PIR, JH0517; JH0517.  
 DR InterPro; IPR000867; Insl\_gro\_fac.pr.  
 DR InterPro; IPR000716; Thyroglobulin\_1.  
 DR PROSITE; PS00222; IGF\_BINDING, PARTIAL.  
 DR PROSITE; PS00484; THYROGLOBULIN\_1; PARTIAL.  
 KW Growth factor binding.  
 FT NON TER 16  
 SQ SEQUENCE 16 AA; 1799 MM; 409888400965552 CRC64;

Query Match 21.9%; Score 25; DB 1; Length 16;  
 Best Local Similarity 55.6%; Pred. No. 1e+03;  
 Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 7 HKIPEEEK 15  
 DB 5 HKPPPEEK 13

## RESULT 7

ID COXB\_THIOB STANDARD; PRT; 20 AA.  
 AC P80974;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Tyrosine c oxidase polypeptide Vb (EC 1.9.3.1) (Fragment).  
 OS Thynnus obesus (Bigeye tuna).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;  
 OC Scombridae; Thunnus.  
 OX NCBI\_TaxID=8241;  
 RN [1]  
 RP SEQUENCE.  
 RT TISSUE=Heart, and Liver;

RP SEQUENCE.  
 RT TISSUE=Heart, and Liver;

RX MEDLINE=97454291; PubMed=9310366;  
 RA Arnold S., Lee I., Kim M., Song E., Linder D., Lottspeich F.,  
 RA Kadenbach B.;  
 RT "The subunit structure of cytochrome-c oxidase from tuna heart and  
 liver.";  
 RL Eur. J. Biochem. 248:99-103(1997).  
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE  
 CC CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN  
 CC MITOCHONDRIAL ELECTRON TRANSPORT.  
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome  
 CC c + 2 H(2)O.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VB FAMILY.  
 DR InterPro; IPR002124; COX5B.  
 DR PROSITE; PS00848; COX5B; PARTIAL.  
 KW Oxidoreductase; Inner membrane; Mitochondrion.  
 FT NON TER 20  
 SQ SEQUENCE 20 AA; 2158 MM; 76F1473E1F392BD7 CRC64;

Query Match 21.9%; Score 25; DB 1; Length 20;  
 Best Local Similarity 33.3%; Pred. No. 1.3e+03;  
 Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 8 KKIPEEKREK 22  
 DB 4 KGIPTDEQATGLER 18

## RESULT 8

ID MOTI\_CANFA STANDARD; PRT; 22 AA.  
 AC P19863;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Motilin.  
 GN Motilin.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]

RP SEQUENCE.  
 RC TISSUE=Intestine;  
 RX MEDLINE=83195948; PubMed=684463;  
 RA Poltias P., Reeve J.R., Jr., Hunkapiller M.W., Hood L.E., Walsh J.H.;  
 RT "Purification and characterization of canine intestinal motilin.";  
 RL Regul. Pept. 5:197-208(1983).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE REGULATION OF  
 CC INTERDIGESTIVE GASTROINTESTINAL MOTILITY AND INDIRECTLY CAUSES  
 CC RHYTHMIC CONTRACTION OF DUODENAL AND COLONIC SMOOTH MUSCLE.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE MOTILIN FAMILY.  
 CC PIR, S00189; S00189.  
 DR PIR, A60313; A60313.  
 KW Hormone.  
 FT UNSURE 1  
 SQ SEQUENCE 22 AA; 2685 MM; 4BECB840ABE0639F CRC64;

Query Match 21.5%; Score 24.5; DB 1; Length 22;  
 Best Local Similarity 37.5%; Pred. No. 1.6e+03;  
 Matches 6; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

OY 5 FAH--KKIPEEKRE 17  
 DB 5 FTTHSELDKIRKERNK 20

## RESULT 9

ID UP41\_UP4IN STANDARD; PRT; 17 AA.  
 AC P82035;  
 DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Uperin 4.1.  
 OS Uperoleia inundata (Floodplain toadlet).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Myobatrachidae;  
 OC Uperoleia.  
 NCBI\_TaxID=104953;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Skin secretion;  
 RA Bradford A.M., Rattery M.J., Bowie J.H., Tyler M.J., Wallace J.C.,  
 Adams G.W., Severini C.;  
 RT "Novel uperin peptides from the dorsal glands of the Australian  
 floodplain toadlet Uperoleia inundata.";  
 RL Aust. J. Chem. 49:475-484 (1996).  
 CC -1- FUNCTION: SHOWS A MEDIUM ANTIBACTERIAL ACTIVITY AGAINST  
 L. MENSETERIODES.  
 CC -1- MASS SPECTROMETRY: MW=1724; METHOD=FAB.  
 KW Amphibian skin; Antibiotic; Amidation.  
 FT MOD RES 17 17  
 SO SEQUENCE 17 AA; 1726 MW; 62E868714BC54D7B CRC64;

Query Match 21.1%; Score 24; DB 1; Length 17;  
 Best Local Similarity 44.4%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGFPAHKKI 10  
 DB 2 VGSFHKV 10

RESULT 10  
 LUXB\_KRYAS STANDARD; PRT; 18 AA.  
 AC P18300;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Alkalal monooxygenase beta chain (EC 1.14.14.3) (Bacterial luciferase  
 beta chain) (Fragment).  
 GN LUXB.  
 OS Kryptophanon alfredi symbiont.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae;  
 OC light emitting symbionts of fish.  
 NCBI\_TaxID=28177;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91076680; PubMed=2256783;  
 RA Haygood M.G.;  
 RT "Relationship of the luminous bacterial symbiont of the Caribbean  
 flashlight fish, Kryptophanon alfredi (family Anamolopidae) to  
 other luminous bacteria based on bacterial luciferase (luxA) genes.";  
 RL Arch. Microbiol. 154:496-503 (1990).  
 CC -1- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA. THE  
 SPECIFIC ROLE OF THE BETA SUBUNIT IS UNKNOWN, BUT IT IS ABSOLUTELY  
 REQUIRED FOR BIOLUMINESCENCE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: RCHO + FMN(2) + O(2) = RCOOH + FMN + H(2)O +  
 light.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, M36597; AAA91214.1; -  
 DR PIR, B44995; B44995.  
 DR InterPro, IPR002103; Bac\_Luciferase.  
 DR PROSITE, PS00494; BACTERIAL\_LUCIFERASE; PARTIAL.

KW Photoprotein; Luminescence; Oxidoreductase; Monooxygenase;  
 KW Flavoprotein; FMN.  
 FT NON TER 18  
 SQ SEQUENCE 18 AA; 2153 MW; 8CB3B6955CCB2E7A CRC64;

Query Match 21.1%; Score 24; DB 1; Length 18;  
 Best Local Similarity 29.4%; Pred. No. 1.6e+03;  
 Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 KLGFPAHKKIPEEKRE 17  
 DB 2 KFGLEFQNFLENSOSE 18

RESULT 11  
 MOTI\_CHICK STANDARD; PRT; 22 AA.  
 ID MOTI\_CHICK  
 AC Q9PRP6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Motilin.  
 GN MOTL.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Small intestine;  
 RX MEDLINE=96255357; PubMed=8801522;  
 RA De Clercq P., Depoortere I., Macielag M., Vandermeers A.,  
 Vandermeers-Piret M.C., Peeters T.L.;  
 RT "Isolation, sequence, and bioactivity of chicken motilin.";  
 RL Peptides 17:203-208 (1996).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE REGULATION OF  
 CC INTERDIGESTIVE GASTROINTESTINAL MOTILITY AND INDIRECTLY CAUSES  
 CC RHYTHMIC CONTRACTION OF DUODENAL AND COLONIC SMOOTH MUSCLE.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE MOTILIN FAMILY.  
 KW Hormone.  
 SQ SEQUENCE 22 AA; 2686 MW; 6FD6A79B7EB02680 CRC64;

Query Match 21.1%; Score 24; DB 1; Length 22;  
 Best Local Similarity 26.3%; Pred. No. 1.9e+03;  
 Matches 5; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 FFAHKKIPEEKREKLEQ 22  
 DB 4 FFGSDIQKQKERNKQ 22

RESULT 12  
 NUOT\_SOLITU STANDARD; PRT; 22 AA.  
 ID NUOT\_SOLITU  
 AC P80730;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase 16 kDa subunit (EC 1.6.5.3)  
 DE (EC 1.6.99.3) (Complex I-16KD) (CI-16KD) (Fragment).  
 OS Solanum tuberosum (Potato).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 NCBI\_TaxID=4113;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=cv. Bintje; TISSUE=Tuber;  
 RA Herz U., Gschmann L.;  
 RL Submitted (DEC-1996) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY

```

CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE.
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: MATRIX SIDE OF THE MITOCHONDRIAL INNER
CC MEMBRANE.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
FT NON TER
SQ SEQUENCE 22 AA; 2435 MW; D4D64C57673C2613 CRC64;

Query Match 21.1%; Score 24; DB 1; Length 22;
Best Local Similarity 62.5%; Pred. No. 1.9e+03;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 8 KKIPEEK 15
DB 4 KVLSEEEK 11

RESULT 13
IDHP RAT STANDARD; PRT; 13 AA.
AC P56574;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Isocitrate dehydrogenase [NADP], mitochondrial (EC 1.1.1.42)
DE (Oxalocaccinate decarboxylase) (IDH) (NADP+-specific ICDH) (IDP) (ICD-
DE M) (Fragment).
GN IDH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RC STRAIN=Mistral; TISSUE=Heart;
RA Li X.-P., Pleissner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,
RA Jungblut P.R.;
RA Submitted (SEP-1998) to the SWISS-PROT data bank.
RL
CC -1- FUNCTION: PLAYS A ROLE IN INTERMEDIARY METABOLISM AND ENERGY
CC PRODUCTION. IT MAY TIGHTLY ASSOCIATE OR INTERACT WITH THE PYRUVATE
CC DEHYDROGENASE COMPLEX (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Isocitrate + NADP(+) = 2-oxoglutarate + CO(2)
CC + NADH.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC (SPOT P8) IS: 9.0, ITS MW IS: 42 kDa.
CC -1- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLMALATE
CC DEHYDROGENASES FAMILY.
CC InterPro: IPR001804; Isoch.
DR PROSITE: PS00470; IDH_IDMH; PARTIAL.
KM Oxidoreductase; NADP; Glyoxylate bypass; Tricarboxylic acid cycle;
KM Mitochondrion.
FT NON TER
SQ SEQUENCE 13 AA; 1526 MW; 5FB81031723B02C3 CRC64;

Query Match 20.2%; Score 23; DB 1; Length 13;
Best Local Similarity 50.0%; Pred. No. 1.5e+03;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 6 AHKKIPEEK 15
DB 1 AEKRIKVEEK 10

RESULT 14
UN04_PINPS STANDARD; PRT; 15 AA.
AC P81673;
DT 15-JUL-1999 (Rel. 38, Created)

```

```

DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Unknown protein from 2D-page of needles (N143) (Fragment).
OS Pinus pinaster (Maritime pine).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=71647;
RN [1]
RP SEQUENCE.
RC TISSUE=Needle;
RX MEDLINE=99274088; PubMed=10344291;
RA Costa P., Pionneau C., Bauw G., Dubos C., Bahman N., Kremer A.,
RA Frigerio J.-M., Plomion C.;
RT "Separation and characterization of needle and xylem maritime pine
RT proteins.";
RL Electrophoresis 20:1098-1108(1999).
CC -1- INDUCTION: BY WATER STRESS.
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.2, ITS MW IS: 21 kDa.
FT NON TER
FT NON TER
SQ SEQUENCE 15 AA; 1489 MW; CE4D85E9308227A CRC64;

Query Match 20.2%; Score 23; DB 1; Length 15;
Best Local Similarity 83.3%; Pred. No. 1.8e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 KKIPEE 13
DB 1 KKIIEE 6

RESULT 15
PSAF MAIZE STANDARD; PRT; 20 AA.
ID P13193;
AC P13193;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Photosystem I reaction centre subunit III (Light-harvesting complex I
DE 17 kDa protein) (Pst-P) (Fragment).
GN PSAP.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE.
RC STRAIN=cv. N273;
RX MEDLINE=90033290; PubMed=2680596;
RA Anandan S., Vainstein A., Thorner J.P.;
RT "Correlation of some published amino acid sequences for photosystem I
RT polypeptides to a 17 kDa LHCI pigment-protein and to subunits III and
RT IV of the core complex.";
RL FEBS Lett. 256:150-154(1989).
CC -1- FUNCTION: PROBABLY PARTICIPATES IN EFFICIENCY OF ELECTRON TRANSFER
CC FROM PLASTOCYANIN TO P700 (OR CYTOCHROME C553 IN ALGAE AND
CC CYANOBACTERIA). THIS PLASTOCYANIN-DOCKING PROTEIN CONTRIBUTES
CC TO THE SPECIFIC ASSOCIATION OF PLASTOCYANIN TO PS I.
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH LUMENAL SIDE OF THE
CC THYLAKOID MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE PSAF FAMILY.
CC PIR: S06150; S06150.
DR MaizeDB; 69342;
KM Photosynthesis; Photosystem I; Chloroplast; Thylakoid; Membrane.
FT NON TER
FT NON TER
SQ SEQUENCE 20 AA; 2110 MW; F0499F98F618897 CRC64;

Query Match 20.2%; Score 23; DB 1; Length 20;
Best Local Similarity 41.7%; Pred. No. 2.4e+03;
Matches 5; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

Oy 11 PEEKREKLEQ 22  
|:|:|:  
Db 8 PKEKKAFAKXEX 19

Search completed: July 16, 2003, 08:13:46  
Job time : 13 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:07:08 / Search time 27 Seconds  
(without alignments)  
167.890 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114  
Sequence: 1 KLGFAHKRIPEEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 7500

Minimum DB seq length: 0  
Maximum DB seq length: 22

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	27.2	19	13	Q91983
2	30	26.3	13	10	Q39380
3	30	26.3	20	4	Q16070
4	30	26.3	22	4	Q9R503
5	29.5	25.9	20	6	Q9TRV7
6	28	24.6	22	4	Q9UC03
7	27.5	24.1	15	2	Q9R525
8	27	23.7	16	4	Q9UC12
9	27	23.7	17	16	Q9K5N8
10	27	23.7	18	6	Q9TRG8
11	27	23.7	20	4	Q15969
12	27	23.7	22	5	Q44710
13	26	22.8	14	11	Q9J0J5
14	26	22.8	15	2	Q9R545
15	26	22.8	15	4	Q9UBK0
16	26	22.8	15	10	P83137

17	26	22.8	19	4	Q9U0G7	Q9UG7 homo sapien
18	25	21.9	9	12	Q9NZH6	Q92766 canine dist
19	25	21.9	16	4	Q9NZH9	Q9NZ9 homo sapien
20	25	21.9	16	6	Q9TR09	Q9TR09 bos taurus
21	25	21.9	16	10	P83142	P83142 malva parvi
22	25	21.9	18	13	Q9PRR7	Q9PR7 gallus galli
23	25	21.9	19	12	Q65747	Q65747 bluetongue
24	25	21.9	20	2	Q9R4Q0	Q9R4Q0 pseudomonas
25	25	21.9	20	4	Q75318	Q75318 homo sapien
26	25	21.9	20	4	Q76086	Q76086 homo sapien
27	25	21.9	20	4	Q9UC88	Q9UC88 homo sapien
28	25	21.9	21	2	Q9R4T3	Q9R4T3 mycobacteri
29	25	21.9	21	10	Q41559	Q41559 triticum ae
30	24.5	21.5	15	10	Q41526	Q41526 triticum ae
31	24	21.1	13	2	Q55234	Q55234 synechocyst
32	24	21.1	15	6	Q9N0F7	Q9N0F7 canis famli
33	24	21.1	16	4	Q9UD21	Q9UD21 homo sapien
34	24	21.1	16	6	Q9N0F6	Q9N0F6 canis famli
35	24	21.1	16	9	Q38671	Q38671 bacterioph
36	24	21.1	17	4	Q9UC66	Q9UC66 homo sapien
37	24	21.1	17	6	Q9TR21	Q9TR21 sus scrofa
38	24	21.1	17	11	Q9QVS7	Q9QVS7 mus sp. lac
39	24	21.1	18	15	Q73573	Q73573 human immun
40	24	21.1	18	15	Q73583	Q73583 human immun
41	24	21.1	18	15	Q73587	Q73587 human immun
42	24	21.1	18	15	Q73603	Q73603 human immun
43	24	21.1	18	15	Q73605	Q73605 human immun
44	24	21.1	18	15	Q73607	Q73607 human immun
45	24	21.1	18	15	Q73613	Q73613 human immun

#### ALIGNMENTS

##### RESULT 1

ID	Q91983	PRELIMINARY;	PRT;	19 AA.
AC	Q91983;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)			
DE	Myosin heavy chain (Fragment).			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	RX MEDLINE=88040428; PubMed=3671071;			
RA	McCarthy J.G., Heywood S.M.;			
RT	"A long polypyrimidine/polypurine tract induces an altered DNA conformation on the 3' coding region of the adjacent myosin heavy chain gene."			
RT	Nucleic Acids Res. 15:8069-8085(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	McCarthy J.G.;			
RL	Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	RX MEDLINE=89210285; PubMed=2707122;			
RA	Lagrutta A.A., McCarthy J.G., Scherzinger C.A., Heywood S.M.;			
RT	"Identification and developmental expression of a novel embryonic myosin heavy-chain gene in chicken."			
RT	DNA 8:39-50(1989).			
RL	EMBL; X06251; CAA2950.1; -			
DR	EMBL; X06251; CAA2950.1; -			
KW	Myosin.			
FT	NON_TER	1	1	
SO	SEQUENCE	19 AA;	2387 MW;	2DD5C6816CD4D6EC CRC64;
Query Match		27.2%;	Score 31;	DB 13; Length 19;

Best Local Similarity 70.0%; Pred. No. 1e+03;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 FAHKKIPPEE 14  
DB 10 FHSKKIEEE 19

## RESULT 2

Q39380 PRELIMINARY; PRT; 13 AA.

AC Q39380; 01-NOV-1996 (TRENBLREL. 01, Created)  
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
DE IFA binding protein (Sp10) (Fragment).  
OS Brassica oleracea (Cauliflower).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC euroside II; Brassicales; Brassicaceae; Brassica.  
OX NCBI\_Taxid=3712;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=DOK; TISSUE=CURD SURFACE;  
RA Willis G.;  
RT "An investigation of nuclear lamin homologues in plants: an apparently non-intermediate filament sequence that bind a polyclonal anti-lamin antiserum."  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X97678; CAA66268.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 13 AA; 1413 MW; DID4EA3926B42772 CRC64;

Query Match 26.3%; Score 30; DB 10; Length 13;  
Best Local Similarity 62.5%; Pred. No. 9.9e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 AHKKIPEE 13  
DB 2 AEKKVPKE 9

Q16070 PRELIMINARY; PRT; 20 AA.  
AC Q16070;  
DT 01-NOV-1996 (TRENBLREL. 01, Created)  
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)  
DE Neurofilament heavy subunit (Fragment).  
GN NEFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94040777; PubMed=8224877;  
RA Figlewicz D.A., Rouleau G.A., Krizus A., Julien J.P.;  
RT "Polymorphism in the multi-phosphorylation domain of the human neurofilament heavy-subunit-encoding gene."  
RL Gene 132:297-300(1993).  
DR EMBL; S66488; AAB28609.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 20 AA; 2198 MW; E9A0975B41FD8082 CRC64;

Query Match 26.3%; Score 30; DB 4; Length 20;  
Best Local Similarity 53.8%; Pred. No. 1.5e+03;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 9 KIPEEKREKLE 21  
DB 6 KSPERAKSPEKEE 18

## RESULT 4

Q9RSU3 PRELIMINARY; PRT; 22 AA.

AC Q9RSU3; 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
DE 2-amino-benzoate-CoA ligase-E2 (Fragment).  
OS Pseudomonas.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae.  
OX NCBI\_Taxid=286;  
RN (1)  
RP SEQUENCE.  
RX MEDLINE=91358327; PubMed=1865526;  
RA Altmerschmidt U., Oswald B., Fuchs G.;  
RT "Purification and characterization of benzoate-coenzyme A ligase and 2-amino-benzoate-coenzyme A ligases from a dentrifying Pseudomonas sp."  
RL J. Bacteriol. 173:5494-5501(1991).  
FT NON\_TER 1  
FT NON\_TER 22  
SQ SEQUENCE 22 AA; 2526 MW; 5D37FA76363F8792 CRC64;

Query Match 26.3%; Score 30; DB 2; Length 22;  
Best Local Similarity 37.5%; Pred. No. 1.6e+03;  
Matches 6; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 5 FAHKKIPPEEKREKL 20  
DB 7 FARDXPPTPEQTESL 22

## RESULT 5

Q9TRY7 PRELIMINARY; PRT; 20 AA.

AC Q9TRY7; 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
DE Insulin-like growth factor-binding protein-5, IGFBP-5 (Fragment).  
OS Sus sp.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
OX NCBI\_Taxid=9826;  
RN (1)  
RP SEQUENCE.  
RX MEDLINE=92049376; PubMed=1719383;  
RA Shimasaki S., Gao L., Shimomura M., Ling N.;  
RT "Isolation and molecular cloning of insulin-like growth factor-binding protein-6."  
RL Mol. Endocrinol. 5:938-948(1991).  
FT NON\_TER 1  
FT NON\_TER 20  
SQ SEQUENCE 20 AA; 2185 MW; F77C34A8FA94EF9 CRC64;

Query Match 25.9%; Score 29.5; DB 6; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.7e+03;  
Matches 7; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LGFFAHKKIPEEK 15  
DB 1 LGSFVHKE-PYDER 13

## RESULT 6

Q9UCO3 PRELIMINARY; PRT; 22 AA.

AC Q9UCO3; 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)

DE KU antigen 73 kDa protein fraction 31 (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Primates; Carnivora; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=92165607; PubMed=1537839;  
RA Medychowski A., Henzel W., Huston L., Paslidis N., Ellerson D.,  
RA McRae M., Seong D., Howard O.M., Deiseroth A.;  
RT "Identification of proteins binding to interferon-inducible  
RT transcriptional enhancers in hematopoietic cells."  
RL J. Biol. Chem. 267:4533-4540 (1992).  
FT NON\_TER 1  
FT NON\_TER 22  
SQ SEQUENCE 22 AA; 2488 MW; 24E2B59F08AF2996 CRC64;

Query Match 24.6%; Score 28; DB 4; Length 22;  
Best Local Similarity 66.7%; Pred. No. 3.1e+03;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 13 EKKREKLE 21  
Db 8 EEDDENLE 16

RESULT 7  
O9R5Z5 PRELIMINARY; PRT; 15 AA.  
AC O9R5Z5;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE Streptolysin O (Fragment).  
OS Streptococcus equisimilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=119602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93273535; PubMed=8500915;  
RA Gerlach D., Kohler W., Gunther E., Mann K.;  
RT "Purification and characterization of streptolysin O secreted by  
RT Streptococcus equisimilis (group C)."  
RL Infect. Immun. 61:2727-2731 (1993).  
SQ SEQUENCE 15 AA; 1716 MW; 7C02547F2FF85BE CRC64;

Query Match 24.1%; Score 27.5; DB 2; Length 15;  
Best Local Similarity 46.7%; Pred. No. 2.5e+03;  
Matches 7; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

OY 6 AHKRIKPEEKREK 19  
Db 1 AKPEKPLESAKEK 15

RESULT 8  
O9UC12 PRELIMINARY; PRT; 16 AA.  
AC O9UC12;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Tropomyosin (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Primates; Carnivora; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=93195352; PubMed=8450225;  
RA Das K.M., Dasgupta A., Mandal A., Geng X.;  
RT "Autoimmunity to cytoskeletal protein tropomyosin. A clue to the

RT pathogenetic mechanism for ulcerative colitis."  
RL J. Immunol. 150:2487-2493 (1993).  
DR InterPro: IPR000533; Tropomyosin.  
DR Pfam: PF00261; Tropomyosin.1.  
SQ SEQUENCE 16 AA; 1960 MW; 7A26C49A1E0A335E CRC64;

Query Match 23.7%; Score 27; DB 4; Length 16;  
Best Local Similarity 62.5%; Pred. No. 3.2e+03;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 14 EKKREKLE 21  
Db 2 QKDEKME 9

RESULT 9  
O9K9N8 PRELIMINARY; PRT; 17 AA.  
AC O9K9N8;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein BH2607.  
GN BH2607.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogawara N., Kihara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis."  
RL Nucleic Acids Res. 28:4317-4331 (2000).  
DR EMBL: AP001516; BAB06326.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 17 AA; 2086 MW; 2238BAPDE271681F CRC64;

Query Match 23.7%; Score 27; DB 16; Length 17;  
Best Local Similarity 50.0%; Pred. No. 3.4e+03;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 8 KKIPEEKREKLE 21  
Db 2 KEIPTKMKIEKDE 15

RESULT 10  
O9TRG8 PRELIMINARY; PRT; 18 AA.  
AC O9TRG8;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE Lactoferrin (Fragment).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=93253156; PubMed=8486845;  
RA Shimazaki K., Tanaka T., Kon H., Oota K., Kawaguchi A., Maki Y.,  
RA Sato T.;  
RL J. Dairy Sci. 76:946-955 (1993).  
SQ SEQUENCE 18 AA; 2138 MW; 809196895D52D7CE CRC64;

Query Match 23.7%; Score 27; DB 6; Length 18;

Best Local Similarity 50.0%; Pred. No. 3.6e+03;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 PEEKREE 18  
| | | | |  
Db 11 PEOKKXQ 18

## RESULT 11

015969 PRELIMINARY; PRT; 20 AA.  
AC 015969;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DE 01-JAN-1999 (TREMBLrel. 09, Last annotation update)  
GN Apolipoprotein A-IV (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93138374; PubMed=1487136;  
RA Kambo M.I., Williams E.R., Law J.C., Aston C.E., Bunker C.H.,  
Ferrell R.E., Pollitzer W.S.;  
RT "Molecular basis of a unique African variant (A-IV 5) of human  
apolipoprotein A-IV and its significance in lipid metabolism.";  
RL Genet. Epidemiol. 9:379-388(1992).  
DR EMBL; S53635; AAB50137.1; -.  
KW Lipoprotein.  
FT NON\_TER 1  
SQ SEQUENCE 20 AA; 2557 MW; 03B96D84967D7DB CRC64;

Query Match 23.7%; Score 27; DB 4; Length 20;  
Best Local Similarity 36.4%; Pred. No. 3.9e+03;  
Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 12 EEEKREKLEQ 22  
| | | | |  
Db 4 QEOQOEQOQEQ 14

## RESULT 12

044710 PRELIMINARY; PRT; 22 AA.  
AC 044710;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
GN Tropomyosin I isoform B (Fragment).  
OS Drosophila pseudoobscura (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7237;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX STRAIN=GOLDENDALE 3;  
RA Hamlin M.T.;  
RT Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF039274; AAB96670.1; -.  
DR EMBL; AF039273; AAB96670.1; JOINED.  
DR FlyBase; FBgn0025413; Dpse\Tml.  
DR InterPro; IPR000533; Tropomyosin.  
DR Pfam; PF00261; Tropomyosin; 1.  
FT NON\_TER 1  
SQ SEQUENCE 22 AA; 2611 MW; F86B844608F4175C CRC64;

Query Match 23.7%; Score 27; DB 5; Length 22;  
Best Local Similarity 31.2%; Pred. No. 4.3e+03;

Matches 5; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREKLE 21  
| | | | |  
Db 1 AEKQYKRLQKEVDRL 16

## RESULT 13

09JUS5 PRELIMINARY; PRT; 14 AA.  
AC 09JUS5;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
GN B-Raf protein (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Barnier J.V.;  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RA Barnier J.V., Papin C., Eyche A., Lecocq O.;  
RT "The mouse B-raf gene encodes multiple protein isoforms with tissue-  
specific expression.";  
RL J. Biochem. 270:23381-23389(1995).  
DR EMBL; AJ276308; CAB81556.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 14 AA; 1748 MW; D1E0505C4927F02 CRC64;

Query Match 22.8%; Score 26; DB 11; Length 14;  
Best Local Similarity 33.3%; Pred. No. 3.9e+03;  
Matches 4; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 8 KKIPEEKREK 19  
| | | | |  
Db 2 EKFEVELQDOR 13

## RESULT 14

09R545 PRELIMINARY; PRT; 15 AA.  
AC 09R545;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
GN 30S ribosomal protein (Fragment).  
OS Mycobacterium bovis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetia; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1765;  
RN (1)  
RP SEQUENCE.  
RX MEDLINE=94009653; PubMed=8405418;  
RA Ohara N., Kimura M., Higashi Y., Yamada T.;  
RT Isolation and amino acid sequence of the 30S ribosomal protein S19  
from Mycobacterium bovis BCG.";  
RL FEBS Lett. 331:9-14(1993).  
SQ SEQUENCE 15 AA; 1674 MW; 07A36F018AE355A4 CRC64;

Query Match 22.8%; Score 26; DB 2; Length 15;  
Best Local Similarity 33.3%; Pred. No. 4.1e+03;  
Matches 4; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 8 KKIPEEKREK 19  
| | | | |  
Db 3 KNVPANSRRKAK 14

## RESULT 15

O9UBKO PRELIMINARY; PRT; 15 AA.  
 ID O9UBKO  
 AC O9UBKO:  
 DT 01-MAY-2000 (TRENBLREL. 13, Created)  
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
 DT 01-JUN-2000 (TRENBLREL. 14, Last annotation update)  
 DE Renal tissue-nonspecific alkaline phosphatase (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=93092315; PubMed=1458595;  
 RA Nishihara Y., Hayashi Y., Adachi T., Koyama I., Stigbrand T.,  
 RA Hirano K.,  
 RT "Chemical nature of intestinal-type alkaline phosphatase in human  
 RT kidney.";  
 RL Clin. Chem. 38:2539-2542(1992).  
 SQ SEQUENCE 15 AA; 1931 MW; 9A28FDE13F01F716 CRC64;  
 OY 10 IPEEK 15  
 :||:|  
 Db 2 VPEEK 7  
 Query Match 22.8%; Score 26; DB 4; Length 15;  
 Best Local Similarity 66.7%; Pred. No. 4.1e+03;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Search completed: July 16, 2003, 08:14:20

Job time : 28 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:56:58 ; Search time 35 Seconds  
(without alignments)  
83.758 Million cell updates/sec

Title: US-09-647-544-7  
Perfect score: 114  
Sequence: 1 KIGFPAHKKIPEEKREKLEKQ 22

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 303745

Minimum DB seq length: 0  
Maximum DB seq length: 22

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: A.Geneseq\_101002.\*  
2: /SID32/gcgdata/geneseq/genesqp-emb1/AA1980.DAT.\*  
3: /SID32/gcgdata/geneseq/genesqp-emb1/AA1981.DAT.\*  
4: /SID32/gcgdata/geneseq/genesqp-emb1/AA1982.DAT.\*  
5: /SID32/gcgdata/geneseq/genesqp-emb1/AA1983.DAT.\*  
6: /SID32/gcgdata/geneseq/genesqp-emb1/AA1984.DAT.\*  
7: /SID32/gcgdata/geneseq/genesqp-emb1/AA1985.DAT.\*  
8: /SID32/gcgdata/geneseq/genesqp-emb1/AA1986.DAT.\*  
9: /SID32/gcgdata/geneseq/genesqp-emb1/AA1987.DAT.\*  
10: /SID32/gcgdata/geneseq/genesqp-emb1/AA1988.DAT.\*  
11: /SID32/gcgdata/geneseq/genesqp-emb1/AA1989.DAT.\*  
12: /SID32/gcgdata/geneseq/genesqp-emb1/AA1990.DAT.\*  
13: /SID32/gcgdata/geneseq/genesqp-emb1/AA1991.DAT.\*  
14: /SID32/gcgdata/geneseq/genesqp-emb1/AA1992.DAT.\*  
15: /SID32/gcgdata/geneseq/genesqp-emb1/AA1993.DAT.\*  
16: /SID32/gcgdata/geneseq/genesqp-emb1/AA1994.DAT.\*  
17: /SID32/gcgdata/geneseq/genesqp-emb1/AA1995.DAT.\*  
18: /SID32/gcgdata/geneseq/genesqp-emb1/AA1996.DAT.\*  
19: /SID32/gcgdata/geneseq/genesqp-emb1/AA1997.DAT.\*  
20: /SID32/gcgdata/geneseq/genesqp-emb1/AA1998.DAT.\*  
21: /SID32/gcgdata/geneseq/genesqp-emb1/AA1999.DAT.\*  
22: /SID32/gcgdata/geneseq/genesqp-emb1/AA2000.DAT.\*  
23: /SID32/gcgdata/geneseq/genesqp-emb1/AA2001.DAT.\*  
24: /SID32/gcgdata/geneseq/genesqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	22	AA132244	Human integrin sub
2	44	38.6	20	AA15156	Alpha IIB beta 3
3	43	37.7	20	AA188078	Integrin alpha-1Ib
4	43	37.7	20	AA159147	Alpha IIB beta 3
5	42	36.8	19	AA148683	Amino acid sequenc
6	39	34.2	20	AA185587	Lung tumour protei
7	38	33.3	22	AA1820986	Protein #2965 enco
8	38	33.3	22	AA168751	Human bone marrow
9	38	33.3	22	AA116576	Peptide #3010 enco
10	37	32.5	20	AA185588	Lung tumour protei

11	36	31.6	13	AA189530	Human Hsp70B' anti
12	36	31.6	13	AA1874746	Transcription fact
13	36	31.6	14	AA1859149	Alpha IIB beta 3
14	36	31.6	16	AA189524	Human Hsp70B' anti
15	36	31.6	16	AA1895148	Alpha IIB beta 3
16	36	31.6	17	AA189533	Hsp70B' immunogeni
17	36	31.6	18	AA152610	v-myb encoded onco
18	36	31.6	19	AA14811	myb oncogene prote
19	36	31.6	19	AA189965	Clan12 B-cell epit
20	36	31.6	19	AA1899534	Hsp70B' immunogeni
21	36	31.6	21	AA1873887	HIV-1 envelope pol
22	36	31.6	21	AA1864349	DP-178 homologue 1
23	36	31.6	21	AA1852640	T20/DP178 peptide
24	36	31.6	21	AA184624	HIV-1 isolate LAI
25	36	31.6	21	AA184624	HIV-1 isolate LAI
26	36	31.6	22	AA184887	Anti-HIV peptide D
27	36	31.6	22	AA182835	Whooping cough ant
28	36	31.6	22	AA184330	DP-178 homologue 2
29	36	31.6	22	AA1852641	T20/DP178 peptide
30	36	31.6	22	AA184625	HIV-1 isolate LAI
31	35	30.7	10	AA184888	Anti-HIV peptide D
32	35	30.7	10	AA184888	Anti-HIV peptide D
33	35	30.7	10	AA184888	Anti-HIV peptide D
34	34	29.8	14	AA187596	Cytoplasmic integr
35	34	29.8	18	AA186614	Human peptide #871
36	34	29.8	19	AA186614	HIV-1 matrix prote
37	34	29.8	20	AA186614	Acidic acid ext
38	34	29.8	20	AA186614	HML-1 alpha-E chat
39	34	29.8	21	AA186614	PEST sequence pep
40	34	29.8	21	AA186614	HML-1 alpha-E chat
41	33	28.9	14	AA186614	Synthetic ubiqit1
42	33	28.9	15	AA186614	Alpha IIB beta 3
43	33	28.9	15	AA186614	Maspin epitope. H
44	33	28.9	19	AA186614	HML-1 alpha-E chat
45	33	28.9	20	AA186614	Fragment of human
			21	AA186614	Human Fas peptide
			21	AA186614	Human Fas epitope,

#### ALIGNMENTS

RESULT 1  
AA132244  
ID AA132244 standard; Peptide: 22 AA.  
XX  
AC AA132244;  
XX  
DT 15-FEB-2000 (first entry)  
XX  
DE Human integrin subunit alpha-10 cytoplasmic domain peptide.  
XX  
KW Integrin alpha-10; ISA10; human; trauma; rheumatoid arthritis;  
KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
KW inflammation; therapy; cartilage; chondrocyte; osteoblast;  
KW Fibroblast; vaccine; marker.  
XX  
OS Homo sapiens.  
XX  
PN WO951639-A1.  
XX  
PD 14-OCT-1999.  
XX  
PF 31-MAR-1999; 99WO-SE00544.  
XX  
PR 02-APR-1998; 98SE-0001164.  
PR 28-JAN-1999; 99SE-0000319.  
XX  
PA (ACT1-) ACTIVE BIOTECH AB.  
XX  
PI Lundgren-Akerlund E.  
XX  
DR WPI; 2000-052639/04.  
XX  
PT New isolated integrin subunit alpha-10, used as a marker or target

PT molecule for cells during development, regeneration and pathological  
PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
PT inflammation

PS Claim 21; Page 53; 90pp; English.

CC This sequence represents a fragment of novel human chondrocyte  
CC integrin subunit alpha-10 (Isa10, see AAY32242), corresponding to  
CC the C-terminal cytoplasmic domain of the protein. The invention  
CC relates to a recombinant or isolated integrin heterodimer  
CC comprising the alpha10 subunit in association with subunit beta  
CC (especially beta-1). The integrin heterodimer, or the subunit  
CC alpha-10, or a fragment of it such as the present sequence, can be  
CC used as a marker or target of all types of cells, e.g. of  
CC chondrocytes, osteoblasts and fibroblasts. They can also be used:  
CC for treating pathological conditions involving Isa10, such as  
CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
CC for detecting the formation of cartilage during embryonal  
CC development, physiological or therapeutic repair of cartilage,  
CC or detecting regeneration of cartilage or chondrocytes during  
CC transplantation of cartilage or chondrocytes; for selection and  
CC analysis or for sorting, isolating or purification of chondrocytes  
CC and for in vitro studies of differentiation of chondrocytes; and as  
CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
CC skeletal muscle or other tissues where adhesion impairs the function  
CC of the tissue (all claimed). Isa10 binding entities can be used to  
CC determine the differentiation-state of cells during embryonic  
CC development, angiogenesis or development of cancer, in pathological  
CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,  
CC in tissue regeneration or in therapeutic and physiological repair  
CC of cartilage (claimed). A vaccine comprising the integrin  
CC heterodimer or subunit alpha-10 is also claimed. Isa10  
CC polynucleotides, vectors, host cells and methods of producing  
CC recombinant Isa10 are also claimed.

XX SQ Sequence 22 AA;

Query Match 100.0%; Score 114; DB 21; Length 22;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLGFPAHKKIPPEEKREKLEQ 22  
DB 1 KLGFPAHKKIPPEEKREKLEQ 22

RESULT 2  
AAB59156  
ID AAB59156 standard; peptide; 20 AA.

XX AC AAB59156;

DT 21-MAR-2001 (first entry)

XX DE Alpha IIB beta 3 binding motif peptide #10.

XX KW Hepatid repeat; transmembrane domain; cytoplasmic; integrin;  
XX inflammation; thrombosis; malignancy.

XX OS Synthetic.

XX PN WO200073341-A1.

XX PD 07-DEC-2000.

XX PF 26-MAY-2000; 2000WO-US14656.

XX PR 27-MAY-1999; 99US-0320907.

XX PA (SCRI ) SCRIPPS RES INST.

XX PI Ginsberg MH, Pfaff M;

XX XX

DR WPI; 2001-041143/05.

XX Polypeptides useful in construction of structural models for:  
PT identifying therapeutic compounds, comprises series of heptad repeats  
PT that mimic a transmembrane domain and cytoplasmic domain attached to  
PT the repeats

PS Disclosure; Page 13; 36pp; English.

CC The present invention relates to a peptide with a series of  
CC heptad-repeats that mimic a transmembrane domain and a selected  
CC cytoplasmic domain attached to the heptad repeats. The invention  
CC is useful for evaluating structure and activity of a selected  
CC occupied and clustered transmembrane protein with the selected  
CC cytoplasmic domain and for identifying therapeutic compounds. It  
CC is also useful for identifying a cytoplasmic domain binding partner.  
CC It is may be used to study protein interactions with transmembrane  
CC proteins such as integrin, which can be used to treat conditions in  
CC which over activity of integrins is involved, such as inflammation,  
CC thrombosis and malignancy.

XX SQ Sequence 20 AA;

Query Match 38.6%; Score 44; DB 22; Length 20;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 KLGFPAHKKIPPEEKRE 18  
DB 1 KVGFPRNRPPEEAAB 18

RESULT 3

ID AAR88078 standard; peptide; 20 AA.

XX AC AAR88078;

DT 30-JUL-1996 (first entry)

XX DE Integrin alpha-IIB cytoplasmic region.

XX KW Helical; soluble; amphiphilic; integrin; chimera; inflammation;  
XX inhibition; thrombosis; malignancy; transmembrane receptor.

XX OS Homo sapiens.

XX PN WO9534641-A1.

XX PD 21-DEC-1995.

XX PE 13-JUN-1995; 95WO-US07542.

XX PR 15-JUN-1994; 94US-0260514.

XX PA (SCRI ) SCRIPPS RES INST.

XX PI Chen Y, Ginsberg MH, Kent SBH, Muir TW, O'Toole TE;

XX PI Williams MJ;

XX DR WPI; 1996-049670/05.

XX New soluble protein models of transmembrane receptors - have helical  
PT amphiphilic sequence to represent transmembrane domain, also  
PT chimeric integrin proteins for inhibiting human cellular integrins  
XX Disclosure; Fig 9; 10pp; English.

CC AAR88078 represents the human integrin alpha-IIB cytoplasmic domain.  
CC It is used to produce a Tac/integrin chimera having the cytoplasmic  
CC domain as above covalently linked to the extracellular and transmembrane  
CC domain of the Tac subunit of the human interleukin-2 receptor. Chimeric  
CC integrins can be used to block the activity of natural integrins in





DT 21-MAY-2002 (first entry)  
XX  
DE Lung tumour protein L548S peptide #22.  
XX  
KW Lung tumour; cancer; T cell; immune response stimulator;  
KW cytostatic.  
XX  
OS Homo sapiens;  
XX  
PN WO200204514-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 10-JUL-2001; 2001WO-US22058.  
XX  
PR 11-JUL-2000; 2000US-0614124.  
PR 29-AUG-2000; 2000US-0651563.  
PR 08-SEP-2000; 2000US-0658824.  
PR 26-SEP-2000; 2000US-0671325.  
PR 06-OCT-2000; 2000US-0677419.  
PR 30-OCT-2000; 2000US-0702705.  
PR 13-DEC-2000; 2000US-0736457.  
PR 03-MAY-2001; 2001US-0849626.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Wang T, Matanabe Y, Henderson RA, Johnson JC, Retter MW;  
PI Marnerakis M, Carter D, Fanger GR, Vedrick TS, Bangur CS;  
PI McNabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;  
XX  
DR WPI; 2002-164634/21.  
XX  
PT Novel polynucleotide encoding a lung tumour polypeptide useful for  
PT stimulating and/or expanding T cells specific for a tumour protein -  
XX  
PS Claim 2; SEQ ID No 1855; 223bp; English.  
XX  
SQ The invention describes an isolated polynucleotide and polypeptide  
CC useful for stimulating and/or expanding T cells specific for a tumour  
CC protein for determining the presence of a cancer in a patient. A  
CC composition containing the polynucleotide and/or polypeptide is useful  
CC for treating a lung cancer in a patient. The polypeptide is useful for  
CC removing tumour cells from a biological sample. The polynucleotide is  
CC also useful as probe or primer to detect the level of mRNA encoding a  
CC tumour protein. This is the amino acid sequence of a lung tumour  
CC associated peptide, described in the method of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 20 AA;  
XX  
Query Match 34.2%; Score 39; DB 23; Length 20;  
Best Local Similarity 53.8%; Pred. No. 81;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 6 AHKKIPEEKREE 18  
DB 8 ARKKVEEDEDEE 20  
XX  
RESULT 7  
ABR20986  
ID ABR20986 standard; Protein; 22 AA.  
XX  
AC ABR20986;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #2985 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;

KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
PS Claim 15; SEQ ID No 22756; 530bp; English.  
XX  
SQ The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 22 AA;  
XX  
Query Match 33.3%; Score 38; DB 22; Length 22;  
Best Local Similarity 61.5%; Pred. No. 1.2e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 7 HKKIPEEKREEK 19  
DB 3 NKKKEEEEEEEK 15  
XX  
RESULT 8  
AAM68751  
ID AAM68751 standard; Protein; 22 AA.  
XX  
AC AAM68751;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29057.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX

30-JAN-2001; 2001WO-US00668.  
04-FEB-2000; 2000US-0180312.  
26-MAY-2000; 2000US-0207456.  
30-JUN-2000; 2000US-0608408.  
03-AUG-2000; 2000US-0632366.  
21-SEP-2000; 2000US-0234687.  
27-SEP-2000; 2000US-0236359.  
04-OCT-2000; 2000GB-0024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI, 2001-488900/53.  
Human genome-derived single exon nucleic acid probes useful for  
analyzing gene expression in human bone marrow -  
Example 4: SEQ ID NO: 29057; 658bp + Sequence Listing; English.  
The present invention provides a number of single exon nucleic acid  
probes which are derived from genomic sequences expressed in the human  
bone marrow. They can be used to measure gene expression in bone marrow  
samples, which may enable the improved diagnosis and treatment of cancers  
such as lymphoma, leukemia and myeloma. The present sequence is a  
protein encoded by one of the probes of the invention.

Sequence 22 AA;

Query Match 33.3%; Score 38; DB 22; Length 22;  
Best Local Similarity 61.5%; Pred. No. 1.2e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 7 HKKIPEEKREK 19  
: || ||| : |||  
Db 3 NKKKEEEEEEEK 15

RESULT 9

AAM16576  
ID AAM16576 standard; Protein; 22 AA.

XX AAM16576;

DT 12-OCT-2001 (first entry)

DE Peptide #3010 encoded by probe for measuring cervical gene expression.

KW Probe; human; microarray; gene expression; cervical epithelial cell;

KV cervical cancer.

XX Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI, 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for  
analyzing gene expression in human cervical epithelial cells -  
Claim 27, SEQ ID No 21402; 487bp; English.  
The present invention relates to human single exon nucleic acid probes  
(SENP: see A110068-A128459). The present sequence is a peptide encoded  
by one such probe. The SENPs are derived from human Hela cells. The SENPs  
can be used to produce a single exon microarray, which can be used for  
measuring human gene expression in a sample derived from human cervical  
epithelial cells. By measuring gene expression, the probes are therefore  
useful in grading and/or staging of diseases of the cervix, notably  
cervical cancer.  
Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 22 AA;

Query Match 33.3%; Score 38; DB 22; Length 22;  
Best Local Similarity 61.5%; Pred. No. 1.2e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 7 HKKIPEEKREK 19  
: || ||| : |||  
Db 3 NKKKEEEEEEEK 15

RESULT 10

AAU85588  
ID AAU85588 standard; Peptide; 20 AA.

AC AAU85588;

DT 21-MAY-2002 (first entry)

DE Lung tumour protein L548S peptide #23.

KW Lung tumour; cancer; T cell; immune response stimulator;

KV cytostatic.

XX Homo sapiens.

PN WO200204514-A2.

PD 17-JAN-2002.

PF 10-JUL-2001; 2001WO-US22058.

PR 11-JUL-2000; 2000US-0614124.

PR 29-AUG-2000; 2000US-0651563.

PR 08-SEP-2000; 2000US-0658824.

PR 26-SEP-2000; 2000US-0671325.

PR 06-OCT-2000; 2000US-0677419.

PR 30-OCT-2000; 2000US-0702705.

PR 13-DEC-2000; 2000US-0736457.

PR 03-MAY-2001; 2001US-0849626.

(CORI-) CORIXA CORP.

Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;

Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;

McNabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;

WPI, 2002-164634/21.

Novel polynucleotide encoding a lung tumour polypeptide useful for  
stimulating and/or expanding T cells specific for a tumour protein -  
Claim 2, SEQ ID No 1856; 223bp; English.  
The invention describes an isolated polynucleotide and polypeptide



CC produced are useful for inhibiting the progression of neoplastic  
 CC diseases. The invention relates to the field of gene therapy and is  
 CC directed toward methods for producing peptide-lipid-polynucleotide  
 CC complexes suitable for delivery of polynucleotides. The encapsulated  
 CC molecules display therapeutic efficacy in eradicating solid tumours  
 CC including but not limited to breast carcinoma or prostate carcinoma.  
 CC ABB74235 co ABB74255 are used in the exemplification of the present  
 CC invention.

CC Sequence 13 AA;

Query Match 31.6%; Score 36; DB 23; Length 13;  
 Best Local Similarity 63.6%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 11 PEEKREKLE 21  
 ||:||||:|  
 Db 3 PEKERIKLE 13

RESULT 13

AAB59149 standard; peptide; 14 AA.

AC AAB59149;

DT 21-MAR-2001 (first entry)

DE Alpha IIB beta 3 binding motif peptide #3.

KW Heptad repeat; transmembrane domain; cytoplasmic; integrin;  
 inflammation; thrombosis; malignancy.

OS Synthetic.

PN WO200073341-A1.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14656.

PR 27-MAY-1999; 99US-0320907.

PA (SCRI) SCRIPPS RES INST.

PI Gineberg MH, Pfaff M;

WPI; 2001-041143/05.

PT Polypeptides useful in construction of structural models for  
 PT identifying therapeutic compounds, comprises series of heptad repeats  
 PT that mimic a transmembrane domain and cytoplasmic domain attached to  
 PT the repeats

PS Disclosure; Page 13; 36pp; English.

CC The present invention relates to a peptide with a series of  
 CC heptad-repeats that mimic a transmembrane domain and a selected  
 CC cytoplasmic domain attached to the heptad repeats. The invention  
 CC is useful for evaluating structure and activity of a selected  
 CC occupied and clustered transmembrane protein with the selected  
 CC cytoplasmic domain and for identifying therapeutic compounds. It  
 CC is also useful for identifying a cytoplasmic domain binding partner.  
 CC It is may be used to study protein interactions with transmembrane  
 CC proteins such as integrin, which can be used to treat conditions in  
 CC which over activity of integrins is involved, such as inflammation,  
 CC thrombosis and malignancy.

CC Sequence 14 AA;

Query Match 31.6%; Score 36; DB 22; Length 14;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 KLGFFPAKKIPEEE 14  
 ||:||||:|  
 Db 1 KVGFFKRRRPPLEE 14

RESULT 14

AAB9524 standard; peptide; 16 AA.

AC AAB9524;

DT 04-SEP-2001 (first entry)

DE Human Hsp70B' antigenic peptide #2.

KW Human; Hsp70B'; detection; stress-inducible protein; antigenic;  
 KW epitope; immunogenic; hypertension; oncology; organ transplantation;  
 KW ischaemia; trauma; infection; inflammation; fever; heart disease;  
 KW autoimmune disorder; neoplasia.

OS Homo sapiens.

PN WO200142423-A2.

PD 14-JUN-2001.

PF 07-DEC-2000; 2000WO-US33341.

PR 07-DEC-1999; 99US-0169535.

PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.

PI Boux HA, Wong GS, Rodriguez H;

WPI; 2001-381658/40.

PT New antibody that specifically binds to heat shock protein 70B' or to  
 PT HSP70B' peptides, useful for determining whether cell has been exposed  
 PT to stressful environment or stressful substance

PS Claim 1; Page 56; 64pp; English.

CC The present invention describes an antibody (I) that specifically binds  
 CC to the human heat shock protein (Hsp70B', or to Hsp70B' peptides. (I)  
 CC is useful for determining whether a cell has been exposed to a stressful  
 CC environment or a stressful substance by an immunoassay in which proteins  
 CC in the cell or proteins extracted from the cell are exposed to (I). (I)  
 CC can be used to distinguish the expression of Hsc70/Hsp70 proteins, which  
 CC occur while a cell is functioning normally and when it is responding to  
 CC stress, and the Hsp70B' protein, which is only produced in response to  
 CC stress. A positive reaction to Hsp70B' antibodies provides an indication  
 CC of the general state of the health of the organism, in which that cell  
 CC resides. The anti-Hsp70B' antibodies can be used to monitor the protein  
 CC levels of a specific, sensitive, native biomarker (Hsp70B') in vitro  
 CC bioassays using human cell lines to evaluate the toxicity of chemical  
 CC compounds. Monitoring Hsp70B' is useful for evaluating the fitness of  
 CC healthy individuals, in the care of patients who are in remission from  
 CC an autoimmune, chronic condition or neoplasia, to evaluate amniotic  
 CC fluids or samples of the placenta and to assess new born infants who are  
 CC at risk. (I) is also useful for evaluation, diagnosis, prognosis and  
 CC continued monitoring of specific disease conditions such as hypertension,  
 CC oncology, organ transplantation, ischaemia, trauma, infection,  
 CC inflammation, fever, heart disease and autoimmune disorders. The present  
 CC sequence represents a specifically claimed Hsp70B' peptide from the  
 CC present invention.

CC Sequence 16 AA;

Query Match 31.6%; Score 36; DB 22; Length 16;  
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 9 KIPPEEKRE 17  
 |||||:  
 Db 3 KIPPEDRK 11

## RESULT 15

AAB59148  
 ID AAB59148 standard; peptide; 16 AA.

AC AAB59148;

DT 21-MAR-2001 (first entry)

DE Alpha IIB beta 3 binding motif peptide #2.

KW Heptad repeat; transmembrane domain; cytoplasmic; integrin;  
 inflammation; thrombosis; malignancy.

OS Synthetic.

PN WO200073341-A1.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14656.

PR 27-MAY-1999; 99US-0320907.

PA (SCRI ) SCRIPPS RES INST.

PI Ginsberg MH, Pfaff M,

DR MPI; 2001-041143/05.

PT Polypeptides useful in construction of structural models for  
 PT identifying therapeutic compounds, comprises series of heptad repeats  
 PT that mimic a transmembrane domain and cytoplasmic domain attached to  
 PT the repeats -

PS Disclosure; Page 13; 36pp; English.

CC The present invention relates to a peptide with a series of  
 CC heptad-repeats that mimic a transmembrane domain and a selected  
 CC cytoplasmic domain attached to the heptad repeats. The invention  
 CC is useful for evaluating structure and activity of a selected  
 CC occupied and clustered transmembrane protein with the selected  
 CC cytoplasmic domain and for identifying therapeutic compounds. It  
 CC is also useful for identifying a cytoplasmic domain binding partner.  
 CC It is used to study protein interactions with transmembrane  
 CC proteins such as integrin, which can be used to treat conditions in  
 CC which over activity of integrins is involved, such as inflammation,  
 CC thrombosis and malignancy.

SQ Sequence 16 AA;

Query Match 31.6%; Score 36; DB 22; Length 16;

Best Local Similarity 50.0%; Pred. No. 1.7e+02;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 KLGFFAKKKIPPEE 14  
 :|||:|  
 Db 1 KVGFFKRNRPPLPE 14

Search completed: July 16, 2003, 08:13:28  
 Job time : 36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:14:28 / Search time 21 Seconds  
(without alignments)  
124.415 Million cell updates/sec

Title: US-09-647-544-7  
Perfect score: 114  
Sequence: 1 KLGFHAKKIPEEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 96116

Minimum DB seq length: 0  
Maximum DB seq length: 22

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PC1\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PC1US\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	38.6	19	US-10-101-487-67	Sequence 67, Appl
2	44	38.6	20	US-09-320-907B-24	Sequence 24, Appl
3	43	37.7	20	US-09-320-907B-15	Sequence 15, Appl
4	39	34.2	20	US-09-736-457-1855	Sequence 1855, Ap
5	39	34.2	20	US-09-902-941-1855	Sequence 1855, Ap
6	39	34.2	20	US-09-849-626-1855	Sequence 1855, Ap
7	39	34.2	20	US-10-017-754-1855	Sequence 1855, Ap
8	38	33.5	22	US-09-864-761-36284	Sequence 36284, A
9	37	33.5	20	US-09-736-457-1856	Sequence 1856, Ap
10	37	33.5	20	US-09-902-941-1856	Sequence 1856, Ap
11	37	33.5	20	US-09-849-626-1856	Sequence 1856, Ap
12	37	33.5	20	US-10-017-754-1856	Sequence 1856, Ap
13	36	31.6	13	US-09-733-179A-5	Sequence 5, Appl
14	36	31.6	13	US-09-876-904A-510	Sequence 510, App
15	36	31.6	14	US-09-320-907B-17	Sequence 17, Appl
16	36	31.6	14	US-09-733-179A-15	Sequence 15, Appl
17	36	31.6	16	US-09-320-907B-16	Sequence 16, Appl
18	36	31.6	16	US-09-733-179A-4	Sequence 4, Appl
19	34	29.8	9	US-10-059-720-4	Sequence 4, Appl

20	34	29.8	20	9	US-09-880-132-1	Sequence 1, Appl
21	34	29.8	20	10	US-09-880-149-1	Sequence 1, Appl
22	34	29.8	20	12	US-10-038-045-15	Sequence 15, Appl
23	34	29.8	21	9	US-10-057-789-231	Sequence 231, App
24	34	29.8	21	9	US-10-212-628-231	Sequence 231, App
25	33.5	29.4	15	9	US-09-320-907B-25	Sequence 25, Appl
26	33.5	29.4	20	9	US-10-142-935-6	Sequence 6, Appl
27	33	28.9	19	9	US-09-984-245-306	Sequence 306, App
28	33	28.9	19	9	US-09-966-262-306	Sequence 306, App
29	33	28.9	19	9	US-09-983-966-306	Sequence 306, App
30	33	28.9	19	9	US-10-143-090-306	Sequence 306, App
31	33	28.9	20	9	US-10-216-484-30	Sequence 30, Appl
32	33	28.9	21	9	US-09-805-301-81	Sequence 81, Appl
33	33	28.1	18	9	US-09-865-989-249	Sequence 249, App
34	32	28.1	18	9	US-10-099-574A-249	Sequence 249, App
35	31	27.2	12	9	US-09-876-904A-300	Sequence 300, App
36	31	27.2	16	9	US-09-832-355A-40	Sequence 40, Appl
37	31	27.2	16	9	US-10-001-073-8	Sequence 8, Appl
38	31	27.2	16	9	US-10-225-567A-948	Sequence 948, App
39	31	27.2	18	9	US-10-116-391-32	Sequence 32, Appl
40	31	27.2	20	9	US-09-945-917-32	Sequence 32, Appl
41	31	27.2	20	10	US-09-777-558-1	Sequence 1, Appl
42	31	27.2	21	9	US-09-974-879-188	Sequence 188, App
43	31	27.2	21	9	US-09-305-736-388	Sequence 388, App
44	30	26.3	13	10	US-09-919-124-26	Sequence 26, Appl
45	30	26.3	13	10	US-09-919-124-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1  
US-10-101-487-67  
Sequence 67, Application US/10101487  
Patent No. US20020169125A1  
GENERAL INFORMATION:  
APPLICANT: LEUNG, DAVID W.  
APPLICANT: BERGMAN, PHILIP A.  
APPLICANT: LOFOUST, ALAN  
APPLICANT: PIETZ, GREGORY E.  
APPLICANT: TOMPKINS, CHRISTOPHER K.  
APPLICANT: WAGNER JR., DAVID W.  
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: 077319/0329  
CURRENT APPLICATION NUMBER: US/10/101,487  
CURRENT FILING DATE: 2002-03-20  
PRIOR APPLICATION NUMBER: 60/277,705  
PRIOR FILING DATE: 2001-03-21  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 67  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide  
US-10-101-487-67  
  
Query Match 38.6%, Score 44; DB 9; Length 19;  
Best Local Similarity 57.1%; Pred. No. 9.5;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
  
CY 9 KIPPEEKREKLEQ 22  
DB 2 KMPPEEEEEEEEE 15  
  
RESULT 2  
US-09-320-907B-24  
Sequence 24, Application US/09320907B  
Publication No. US20020198360A1  
GENERAL INFORMATION:

```

; APPLICANT: Bangur, Chaitany
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: Switzer, Anne

```



```

; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1855
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-1855

Query Match          34.2%; Score 39; DB 9; Length 20;
Best Local Similarity 53.8%; Pred. No. 48;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 6 AHKKIPEEEKKEE 18
    |||:||||:
Db 8 ARKKVEEDEDEEE 20

RESULT 7
US-10-017-754-1855
; Sequence 1855, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedavick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1855
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-754-1855

Query Match          34.2%; Score 39; DB 9; Length 20;
Best Local Similarity 53.8%; Pred. No. 48;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 6 AHKKIPEEEKKEE 18
    |||:||||:
Db 8 ARKKVEEDEDEEE 20

RESULT 8
US-09-864-761-36284
; Sequence 36284, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
```

```

; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36284
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007158.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
US-09-864-761-36284

Query Match          33.3%; Score 38; DB 10; Length 22;
Best Local Similarity 61.5%; Pred. No. 72;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 7 HKKIPEEEKKEE 19
    |||:||||:
Db 3 NKKEEKEEKEE 15

RESULT 9
US-09-736-457-1856
; Sequence 1856, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedavick, Tom
; APPLICANT: Carter, Derrick
; APPLICANT: Retter, Marc
```

```

; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1856
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1856

Query Match      32.5%; Score 37; DB 9; Length 20;
Best Local Similarity 40.0%; Pred. No. 89;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY      8 KKIPEEKREKLEQ 22
DB      2 KKVEEDEDEQEEREE 16
```

```

RESULT 10
US-09-902-941-1856
; Sequence 1856, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1856
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-941-1856
```

```

Query Match      32.5%; Score 37; DB 9; Length 20;
Best Local Similarity 40.0%; Pred. No. 89;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY      8 KKIPEEKREKLEQ 22
DB      2 KKVEEDEDEQEEREE 16
```

```

RESULT 11
US-09-849-626-1856
; Sequence 1856, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tonglong
; APPLICANT: Switzer, Anne
```

```

; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1856
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-1856
```

```

Query Match      32.5%; Score 37; DB 9; Length 20;
Best Local Similarity 40.0%; Pred. No. 89;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
```

```

QY      8 KKIPEEKREKLEQ 22
DB      2 KKVEEDEDEQEEREE 16
```

```

RESULT 12
US-10-017-754-1856
; Sequence 1856, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1856
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-754-1856
```

```

Query Match      32.5%; Score 37; DB 9; Length 20;
Best Local Similarity 40.0%; Pred. No. 89;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
```

```

QY      8 KKIPEEKREKLEQ 22
DB      2 KKVEEDEDEQEEREE 16
```

```

RESULT 13
US-09-733-179A-5
; Sequence 5, Application US/09733179A
; Publication No. US20030073160A1
; GENERAL INFORMATION:
; APPLICANT: Boux, Heather A.
; APPLICANT: Wong, Geraldine S.
; APPLICANT: Rodriguez, Henry
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING STRESS-INDUCIBLE PROTEINS
; FILE REFERENCE: 12071-006001
; CURRENT APPLICATION NUMBER: US/09/733,179A
```

```
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: WO US00/33341
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,535
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-733-179A-5

Query Match
Best Local Similarity 31.6%; Score 36; DB 9; Length 13;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 9 KPEEKRE 17
Db 3 KPEEDRRK 11

RESULT 14
US-09-876-904A-510
; Sequence 510, Application US/09876904A
; Publication No. US20030072794A1
; GENERAL INFORMATION:
; APPLICANT: BOULIKAS, TENI
; TITLE OF INVENTION: ENCAPSULATION OF PLASMID DNA (LIPOMES TM) AND THERAPEUTIC
; TITLE OF INVENTION: AGENTS WITH NUCLEAR LOCALIZATION SIGNAL/FUSOGENIC PEPTIDE
; FILE REFERENCE: TB-2002.00
; CURRENT APPLICATION NUMBER: US/09/876,904A
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,925
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 510
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: Mouse c-Myb
US-09-876-904A-510

Query Match
Best Local Similarity 31.6%; Score 36; DB 9; Length 13;
Best Local Similarity 63.6%; Pred. No. 77;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 PEEKREKLE 21
Db 3 PEKREKLE 13

RESULT 15
US-09-320-907B-17
; Sequence 17, Application US/09320907B
; Publication No. US20020198360A1
; GENERAL INFORMATION:
; APPLICANT: GINSBERG, MARK H.
; APPLICANT: PFAFF, MARTIN
; TITLE OF INVENTION: STRUCTURAL MODELS FOR CYTOPLASMIC DOMAINS OF
; TITLE OF INVENTION: TRANSMEMBRANE RECEPTORS
; FILE REFERENCE: SRI-0006
; CURRENT APPLICATION NUMBER: US/09/320,907B
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 09/187,236
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 14
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-320-907B-17

Query Match
Best Local Similarity 31.6%; Score 36; DB 9; Length 14;
Best Local Similarity 50.0%; Pred. No. 84;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 KLGFEAKKPEEE 14
Db 1 KVGFFKRNRPLEE 14

Search completed: July 16, 2003, 08:21:30
Job time : 22 secs
```



GenCore version 5.1.6  
Copyright (c), 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 16, 2003, 08:26:05 ; Search time 14 Seconds  
(without alignments)  
3457.348 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167  
Sequence: 1 MELPFVTHLFLPLVFLTGIC.....GFFAHKKIPEEKREKLEQ 1167Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1167	100.0	1167	1	ITAG_HUMAN
2	13	1.1	285	1	ITAI_CHICK
3	13	1.1	1151	1	ITAI_HUMAN
4	13	1.1	1180	1	ITAI_HUMAN
5	13	1.1	1189	1	ITAH_HUMAN
6	11	0.9	687	1	CSTRA_HELPI
7	11	0.9	687	1	CSTRA_HELPI
8	11	0.9	1170	1	ITAI_BOVIN
9	11	0.9	1178	1	ITAI_MOUSE
10	11	0.9	1181	1	ITAI_HUMAN
11	10	0.8	1179	1	ITAH_HUMAN
12	9	0.8	1163	1	ITAI_MOUSE
13	9	0.8	1170	1	ITAI_HUMAN
14	8	0.7	76	1	CD24_MOUSE
15	8	0.7	76	1	CD24_MOUSE
16	8	0.7	80	1	CD24_HUMAN
17	8	0.7	108	1	PRVB_MERB1
18	8	0.7	179	1	IL22_HUMAN
19	8	0.7	234	1	EMB2_CAVPO
20	8	0.7	254	1	PSAV_YEAST
21	8	0.7	256	1	HISU_CAMJE
22	8	0.7	259	1	PSA7_SCHPO
23	8	0.7	264	1	COMT_RAT
24	8	0.7	309	1	ALSK_ECOLI
25	8	0.7	423	1	NSMA_HUMAN
26	8	0.7	470	1	CPBK_MOUSE
27	8	0.7	491	1	CPB1_RAT
28	8	0.7	491	1	CPB1_RAT
29	8	0.7	500	1	CPBA_MOUSE
30	8	0.7	604	1	ITAB_PAPCY
31	8	0.7	643	1	ZN74_HUMAN
32	8	0.7	942	1	TMK1_ARATH
33	8	0.7	959	1	VP2_BTIV3V

34	8	0.7	1033	1	ITAB_MOUSE	O9qum0 mus musculus
35	8	0.7	1039	1	ITAB_HUMAN	P08514 homo sapien
36	8	0.7	1050	1	ITAS_XENLA	O06274 xenopus lae
37	8	0.7	1146	1	ITAI_DROME	O24247 drosophila
38	8	0.7	1152	1	ITAM_HUMAN	P11215 homo sapien
39	8	0.7	1163	1	ITRX_HUMAN	P20702 homo sapien
40	8	0.7	3176	1	CA36_HUMAN	P12111 homo sapien
41	8	0.6	72	1	DMS4_AGAN	O93222 agalychnis
42	7	0.6	73	1	DMS2_AGAN	O93222 agalychnis
43	7	0.6	74	1	DMS3_PHYBI	P81485 phyllomedus
44	7	0.6	77	1	ACP_FOMTE	P80918 comamonas t
45	7	0.6	79	1	ACEI_RALSO	O8y0j1 talstonia s
46	7	0.6	79	1	ACEI_XYLA	O9pfi5 xyella tas
47	7	0.6	97	1	YAJC_HAETN	P44352 haemophilus
48	7	0.6	107	1	INS_CHICK	P01332 gallus gall
49	7	0.6	110	1	HDEA_ECOLI	P26604 escherichia
50	7	0.6	139	1	IGF_MYXGL	P22618 myxine glut
51	7	0.6	142	1	CDU_YEAST	O06549 saccharomyc
52	7	0.6	142	1	MR_CHICK	P24052 gallus gall
53	7	0.6	144	1	GLB_APLJU	P14393 aplysia jul
54	7	0.6	149	1	RHIC_RHILV	O03315 rhizobium 1
55	7	0.6	160	1	BIK_HUMAN	O13323 homo sapien
56	7	0.6	170	1	YGR5_MYXXA	O06929 myxococcus
57	7	0.6	171	1	ATPF_SYNE6	P08447 synechococc
58	7	0.6	171	1	YF87_METU	O58982 methanococc
59	7	0.6	178	1	PORC_METU	O57717 methanococc
60	7	0.6	181	1	HYEF_ECOLI	P77423 escherichia
61	7	0.6	198	1	UVP1_ECOLI	P18957 escherichia
62	7	0.6	205	1	RAC1_DICDI	O9gpr2 dictyosteli
63	7	0.6	206	1	GPBB_MOUSE	P56400 mus musculu
64	7	0.6	208	1	R13_SCHPO	O74175 schizosach
65	7	0.6	212	1	RPSD_MYCTU	O50712 mycobacteri
66	7	0.6	214	1	RL40_CABEL	O09533 caenorhabdi
67	7	0.6	220	1	FGP3_CHICK	P48801 gallus gall
68	7	0.6	221	1	SDFL_HUMAN	O9hcn8 homo sapien
69	7	0.6	225	1	EMBP_MOUSE	O61878 mus musculu
70	7	0.6	225	1	BRX1_HUMAN	O9hbu1 homo sapien
71	7	0.6	227	1	EMBP_RAT	O63189 rattus norv
72	7	0.6	231	1	NH40_CABEL	O22127 caenorhabdi
73	7	0.6	231	1	Y667_RICPR	O9gcn9 rickettsia
74	7	0.6	233	1	EMB1_CAVPO	P22032 cavia porce
75	7	0.6	235	1	COMB_CLOAB	O97082 clostridium
76	7	0.6	235	1	YOGA_ECOLI	O46831 escherichia
77	7	0.6	240	1	IBP6_HUMAN	P24592 homo sapien
78	7	0.6	243	1	YCGE_HUMAN	P75989 escherichia
79	7	0.6	251	1	TPIS_STREX	P82478 streptococc
80	7	0.6	251	1	Y4BF_RHISN	P54429 rhizobium s
81	7	0.6	254	1	BRX1_MOUSE	O9er42 mus musculu
82	7	0.6	256	1	KLKF_HUMAN	O9h2t5 homo sapien
83	7	0.6	258	1	KPM1_ECOLI	P23889 escherichia
84	7	0.6	258	1	KPM2_ECOLI	P24584 escherichia
85	7	0.6	260	1	NK3_XENLA	P25435 xenopus lae
86	7	0.6	262	1	SP1B_HUMAN	O01892 homo sapien
87	7	0.6	266	1	SYGB_MORCA	P77815 moraxella c
88	7	0.6	269	1	TRBU_AGRU	P54911 agrobacteri
89	7	0.6	269	1	VI03_VACCC	P20499 vaccinia vi
90	7	0.6	269	1	VI03_VACCV	P12923 vaccinia vi
91	7	0.6	269	1	VI03_VACV	P33000 variola vir
92	7	0.6	270	1	REB8_HUMAN	O00124 homo sapien
93	7	0.6	289	1	ROCI_ARATH	O9hnu4 arabidopsis
94	7	0.6	295	1	PNK_PSEAE	O93204 pseudomonas
95	7	0.6	307	1	PNK2_SYNY3	P73955 synechocyst
96	7	0.6	312	1	CAH4_BOVIN	O95323 bos taurus
97	7	0.6	315	1	YG4W_YEAST	P50085 saccharomyc
98	7	0.6	319	1	Y305_BROME	O8d472 bruceella me
99	7	0.6	321	1	MRAY_ENTHR	O07668 enterococcu
100	7	0.6	324	1	AAIR_CHICK	P49892 gallus gall

## ALIGNMENTS

RESULT 1

ITAG\_HUMAN  
ID ITAG\_HUMAN STANDARD; PRT; 1167 AA.  
AC 075578; Q9UH28;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-10 precursor.  
GN ITGA10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Articular chondrocytes;  
RX MEDLINE=98352078; PubMed=9685391;  
RA Camper L., Hellman U., Lundgren-Akerlund E.;  
RT "Isolation, cloning, and sequence analysis of the integrin subunit  
alpha10, a beta1-associated collagen binding integrin expressed on  
chondrocytes.";  
RT J. Biol. Chem. 273:20383-20389(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart, and Endothelial cells;  
RX MEDLINE=20169197; PubMed=10702680;  
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,  
Wang S.-X., Langley R., Kriessan G.W.;  
RT "The integrin alpha10 subunit: expression pattern, partial gene  
structure, and chromosomal localization.";  
RT Cytogenet. Cell Genet. 87:238-244(1999).  
CC -1- SUBUNIT: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.  
CC -1- FUNCTION: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10  
CC ASSOCIATES WITH BETA-1.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN  
CC MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.  
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WFPA DOMAIN. INTEGRINS  
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 WFPA DOMAIN.  
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF074015; AAC31952.1; -  
CC EMBL; AF112345; AAF21944.1; -  
CC EMBL; AF172723; AAF61638.1; -  
CC HSSP; P17301; 1A0X.  
CC GeneW; HGNC:6135; ITGA10.  
CC MIM; 604042; -  
CC InterPro; IPR000413; Integrin\_alpha.  
CC InterPro; IPR002035; WVF\_A.  
CC Pfam; PF00092; wva; 1.  
CC Pfam; PF01839; FG-GAP; 5.  
CC PRINTS; PR01185; INTEGRIN.  
CC PRINTS; PR00453; WVFADOMAIN.  
CC SMART; SM00191; Int\_alpha; 5.  
CC SMART; SM00327; WVA; 1.  
CC PROSITE; PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.  
CC PROSITE; PS00234; WVF\_A; 1.  
CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
KW Signal; Repeat; Calcium; Magnesium.  
FT SIGNAL 1 22  
FT CHAIN 23 1167  
FT DOMAIN 23 1122  
FT TRANSMEM 1123 1145  
FT DOMAIN 1146 1167  
CYTOPLASMIC (POTENTIAL).

FT REPEAT 38 97  
FT REPEAT ? ?  
FT DOMAIN 167 350  
FT REPEAT 365 427  
FT REPEAT 428 482  
FT REPEAT 483 545  
FT REPEAT 546 605  
FT REPEAT 608 660  
FT DOMAIN 1134 1140  
FT CA\_BIND 494 502  
FT CA\_BIND 558 566  
FT CA\_BIND 620 628  
FT DISULFID 76 86  
FT DISULFID 666 675  
FT DISULFID 681 736  
FT DISULFID 789 795  
FT CARBOHYD 98 98  
FT CARBOHYD 224 234  
FT CARBOHYD 336 336  
FT CARBOHYD 364 364  
FT CARBOHYD 733 733  
FT CARBOHYD 763 763  
FT CARBOHYD 839 839  
FT CARBOHYD 921 921  
FT CARBOHYD 1011 1011  
FT CARBOHYD 1018 1018  
FT CARBOHYD 1039 1039  
FT CONFLICT 844 844  
FT CONFLICT 909 909  
FT CONFLICT 926 926  
SQ SEQUENCE 1167 AA; 127573 MW; AETD3ALC25CIAE0 CRC64;  
Query Match 100.0%; Score 1167; DB 1; Length 1167;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MELPVTHLFLPLVLTGLGSPFNIDHHPRLPFPPEAEFGYSVLQHYGGGQRMALVGA 60  
1 MELPVTHLFLPLVLTGLGSPFNIDHHPRLPFPPEAEFGYSVLQHYGGGQRMALVGA 60  
DB 1 MELPVTHLFLPLVLTGLGSPFNIDHHPRLPFPPEAEFGYSVLQHYGGGQRMALVGA 60  
OY 121 FMAACPLMSRACSSVFSSGICARVDASFQGGSLAPFAQCPFTMDVIVLDGNSIYP 180  
121 FMAACPLMSRACSSVFSSGICARVDASFQGGSLAPFAQCPFTMDVIVLDGNSIYP 180  
DB 121 FMAACPLMSRACSSVFSSGICARVDASFQGGSLAPFAQCPFTMDVIVLDGNSIYP 180  
OY 181 MSEVOTFLRLVGLKFLIPEDQVGLVOYGSPVHNSLGPRTKEEVVRAAKNLSREG 240  
181 MSEVOTFLRLVGLKFLIPEDQVGLVOYGSPVHNSLGPRTKEEVVRAAKNLSREG 240  
DB 181 MSEVOTFLRLVGLKFLIPEDQVGLVOYGSPVHNSLGPRTKEEVVRAAKNLSREG 240  
OY 241 RETKTAQAIMVACTGFSQSHGPREARLLVVVTGDSHSGEELPALKKAQCEARVRY 300  
241 RETKTAQAIMVACTGFSQSHGPREARLLVVVTGDSHSGEELPALKKAQCEARVRY 300  
DB 241 RETKTAQAIMVACTGFSQSHGPREARLLVVVTGDSHSGEELPALKKAQCEARVRY 300  
OY 301 GIAVLGHLRLQRDPSFLREIRITASDPDRFFNFVDEAALTDIVDALGDRIFGLEGS 360  
301 GIAVLGHLRLQRDPSFLREIRITASDPDRFFNFVDEAALTDIVDALGDRIFGLEGS 360  
DB 301 GIAVLGHLRLQRDPSFLREIRITASDPDRFFNFVDEAALTDIVDALGDRIFGLEGS 360  
OY 361 HAENSSFGLEMSQIGESTHRLKDGILFGMGAYDMGSSVLMLEGGHLLFPPRMALDEBF 420  
361 HAENSSFGLEMSQIGESTHRLKDGILFGMGAYDMGSSVLMLEGGHLLFPPRMALDEBF 420  
DB 361 HAENSSFGLEMSQIGESTHRLKDGILFGMGAYDMGSSVLMLEGGHLLFPPRMALDEBF 420  
OY 421 PPALONHAAYGYSSVSWLLRGRLFLPSGAPRFRHRSKVAFOLKKGAARVAQSLDGE 480  
421 PPALONHAAYGYSSVSWLLRGRLFLPSGAPRFRHRSKVAFOLKKGAARVAQSLDGE 480  
DB 421 PPALONHAAYGYSSVSWLLRGRLFLPSGAPRFRHRSKVAFOLKKGAARVAQSLDGE 480  
OY 481 QIGSYFSGSELPLDTRDGTDDVLLVAAPMFLGPONKETGRVYVYLQGOQSLTLQGTLO 540  
481 QIGSYFSGSELPLDTRDGTDDVLLVAAPMFLGPONKETGRVYVYLQGOQSLTLQGTLO 540  
DB 481 QIGSYFSGSELPLDTRDGTDDVLLVAAPMFLGPONKETGRVYVYLQGOQSLTLQGTLO 540  
OY 541 PEPPODARFFGAMGALPLNDGDGADVAVGAPLEDHGOGALYLHGTGSGVRPAPAORIA 600  
541 PEPPODARFFGAMGALPLNDGDGADVAVGAPLEDHGOGALYLHGTGSGVRPAPAORIA 600

```

Db 541 PPPPDARGFAMGALPDLNODGFADVAAGAPLEBCHOGALVLYHGTOSGVPHPAORLA 600
Oy 601 AAMSMHALSYFGSRVDGRDLDDGDDLVDAVAGNOGAILLSRPVYHLPSLEVPQAL 660
Db 601 AAMSMHALSYFGSRVDGRDLDDGDDLVDAVAGNOGAILLSRPVYHLPSLEVPQAL 660
Oy 661 VVORDCRRRGQBAVCLTAALCFQVTSRTPGRMDHGFYKMFATSLDEWTAGARAFDGSQ 720
Db 661 VVORDCRRRGQBAVCLTAALCFQVTSRTPGRMDHGFYKMFATSLDEWTAGARAFDGSQ 720
Oy 721 RLSRRLRLSVGNVTCQHLHFVLDTSVLRPVALTVALNTTKPGVNLNEGSPTSIQ 780
Db 721 RLSRRLRLSVGNVTCQHLHFVLDTSVLRPVALTVALNTTKPGVNLNEGSPTSIQ 780
Oy 781 KLVPFSKDCGPNCECTDLVLQVNDIRSRKAPFVVGGRKRVLVTLENRKNAYNT 840
Db 781 KLVPFSKDCGPNCECTDLVLQVNDIRSRKAPFVVGGRKRVLVTLENRKNAYNT 840
Oy 841 SLSIFSRNLHLASLTPQRESPIKVECAPSAHARLCSGHPVFTGAKVTFLLFEFSC 900
Db 841 SLSIFSRNLHLASLTPQRESPIKVECAPSAHARLCSGHPVFTGAKVTFLLFEFSC 900
Oy 901 SLLSQVRFKLTASSDSLEFRNGTLQENTQTSAYIOYEPHLLFSSSESTLHRYEHPYGT 960
Db 901 SLLSQVRFKLTASSDSLEFRNGTLQENTQTSAYIOYEPHLLFSSSESTLHRYEHPYGT 960
Oy 961 PVGPGPEFPTLRVNLGCVVSGLLISALPVAHAGVYFSLQVITNNNSCIVONT 1020
Db 961 PVGPGPEFPTLRVNLGCVVSGLLISALPVAHAGVYFSLQVITNNNSCIVONT 1020
Oy 1021 EPPGPPVHPEELOHTNRLNGSNTQCVVACHLGOLAKGTEVSGLLRLVHNEFFRRAK 1080
Db 1021 EPPGPPVHPEELOHTNRLNGSNTQCVVACHLGOLAKGTEVSGLLRLVHNEFFRRAK 1080
Oy 1081 SLLTVSTFELGTBESGVLTQTSRWSESLLEVQTRPILISLWILIGSVLGLLALL 1140
Db 1081 SLLTVSTFELGTBESGVLTQTSRWSESLLEVQTRPILISLWILIGSVLGLLALL 1140
Oy 1141 VFCLMKLGFPAHKKIPEEKREKLEQ 1167
Db 1141 VFCLMKLGFPAHKKIPEEKREKLEQ 1167

```

RESULT 2  
ITAL\_CHICK STANDARD; PRT; 285 AA.

AC Q90615; 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a)  
DE (Fragment).  
GN ITGAL.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=izzard;  
RX MEDLINE=94357930; PubMed=7521332;  
RA Kern A., Briesewitz R., Bank I., Marcantonio E.E.;  
RT "The role of the I domain in ligand binding of the human integrin  
alpha 1 beta 1.";  
RL J. Biol. Chem. 269:22811-22816(1994).  
CC -I- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND  
COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-  
E-R IN COLLAGEN.  
CC -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1  
CC ASSOCIATES WITH BETA-1.  
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.

```

CC -I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U10114; AAA59067.1; -.
CC DR HSPB; P17301; IAOX.
CC DR InterPro; IPR000413; Integrin_alpha.
CC DR InterPro; IPR002035; VWFA_A.
CC DR Pfam; PF00092; vwa; 1.
CC DR SMART; SM00327; VWA; 1.
CC DR PROSITE; PS00242; INTEGRIN_ALPHA; PARTIAL.
CC DR PROSITE; PS50234; VWFA; 1.
CC KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane.
CC FT NON_TER 1 1
CC FT DOMAIN 1 285 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 66 279 VWFA.
CC FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT NON_TER 285 285
CC SQ SEQUENCE 285 AA; 31503 MW; 1B05D3246CDSCAYE CRC64;

```

Query Match 1.1%; Score 13; DB 1; Length 285;  
Best Local Similarity 100.0%; Pred. No. 9e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 169 VIVLDGNSNIYPM 181  
Db 65 VIVLDGNSNIYPM 77.

RESULT 3  
ITAL\_HUMAN STANDARD; PRT; 1151 AA.

AC P56159; 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).  
DE ITGAL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93155124; PubMed=8428973;  
RA Briesewitz R., Epstein M.R., Marcantonio E.E.;  
RT "Expression of native and truncated forms of the human integrin alpha  
1 subunit.";  
RL J. Biol. Chem. 268:2989-2996(1993).  
CC -I- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND  
COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-  
E-R IN COLLAGEN.  
CC -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1  
CC ASSOCIATES WITH BETA-1.  
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -I- SIMILARITY: CONTAINS 1 FG-GAP REPEATS.  
CC -I- DATABASE: NAME=PROV; NOTE=CD guide CD49a entry;





```

FT REPEAT ? ? FG-GAP 2.
FT DOMAIN 175 388 VMPA.
FT REPEAT 377 432 FG-GAP 3.
FT REPEAT 433 484 FG-GAP 4.
FT REPEAT 485 565 FG-GAP 5.
FT REPEAT 567 626 FG-GAP 6.
FT REPEAT 629 681 FG-GAP 7.
FT CA_BIND 497 505 POTENTIAL.
FT CA_BIND 579 587 POTENTIAL.
FT CA_BIND 641 649 POTENTIAL.
FT SITE 1168 1172 GPCR MOTIF.
FT DISULFID 82 92 BY SIMILARITY.
FT DISULFID 687 696 BY SIMILARITY.
FT DISULFID 702 755 BY SIMILARITY.
FT DISULFID 807 813 BY SIMILARITY.
FT DISULFID 877 885 BY SIMILARITY.
FT DISULFID 1029 1062 BY SIMILARITY.
FT DISULFID 1066 1073 BY SIMILARITY.
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 459 459 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1007 1007 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1103 1103 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1180 AA; 130808 MW; 855DA2BE02362EB4 CRC64;

Query Match 1.1%; Score 13; DB 1; Length 1180;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RT human integrin alpha11 subunit (ITGA11).";
RL Genomics 60:179-187(1999).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395147; PubMed=10464311.
RA Velling T., Kusche-Gullberg M., Sejersten T., Gullberg D.,
RT "CDNA Cloning and Chromosomal Localization of Human alpha(11)
RT integrin. A collagen-binding, i domain-containing, beta(1)-associated
RT integrin alpha-chain present in muscle tissues.";
RL J. Biol. Chem. 274:25735-25742(1999).
[3]
RP SEQUENCE OF 954-1188 FROM N.A.
RT Tissue=Fibroblast;
RA Andreu N., Escivill X., Escarceller M., Sunmy L.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
CC -! FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC -! SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11
CC -! ASSOCIATES WITH BETA-1.
CC -! SUBCELLULAR LOCATION: Type I membrane protein.
CC -! TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND
CC HEART. INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO
CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO
CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS,
CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN
CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.
CC -! DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING
CC FETAL MUSCLE CELLS (IN VITRO).
CC -! DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VMPA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -! SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -! SIMILARITY: CONTAINS 1 VMPA DOMAIN.
CC -! SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF109681; AAF01258.1; -
DR EMBL; AF137378; AAD51919.2; -
DR EMBL; AL359064; CAB94392.1; -
DR HSSP; P17301; IAOX
DR Genew; HGNC:6136; ITGA11.
DR MIM; 604789; -
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VMP_A.
DR Pfam; PF00092; vwa; 1.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PRO1185; INTEGRIN.
DR PRINTS; PRO0453; VMPADOMAIN.
DR SMART; SMO0191; Int_alpha; 5.
DR SMART; SMO0327; VMA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS50234; VMPA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium; Magnesium; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1189 INTEGRIN_ALPHA-11.
FT DOMAIN 23 1142 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1143 1165 POTENTIAL.
FT DOMAIN 1166 1189 CYTOPLASMIC (POTENTIAL).
FT REPEAT 38 94 FG-GAP 1.
FT REPEAT 102 163 FG-GAP 2.
FT REPEAT 167 345 VMPA.
FT REPEAT 359 420 FG-GAP 3.
FT REPEAT 422 475 FG-GAP 4.
FT REPEAT 477 537 FG-GAP 5.
FT REPEAT 539 598 FG-GAP 6.
FT REPEAT 601 653 FG-GAP 7.

```

```

FT DOMAIN 1154 1162 POLY-LEU.
FT DOMAIN 1174 1177 POLY-ARG.
FT CA_BIND 488 496 POTENTIAL.
FT CA_BIND 551 559 POTENTIAL.
FT CA_BIND 613 621 POTENTIAL.
FT DISULFID 76 83 BY SIMILARITY.
FT DISULFID 121 139 POTENTIAL.
FT DISULFID 129 159 POTENTIAL.
FT DISULFID 659 668 BY SIMILARITY.
FT DISULFID 674 729 BY SIMILARITY.
FT DISULFID 781 787 BY SIMILARITY.
FT DISULFID 881 893 BY SIMILARITY.
FT CARBOHYD 82 82 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 642 642 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 694 694 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1032 1032 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 433 433 V -> M.
FT VARIANT 524 524 /FTID=VAR_009899.
FT VARIANT 524 524 R -> L.
FT VARIANT 972 972 /FTID=VAR_009890.
FT VARIANT 972 972 L -> P.
FT VARIANT 1003 1003 /FTID=VAR_009891.
FT VARIANT 1003 1003 I -> M.
FT VARIANT 1030 1030 /FTID=VAR_009892.
FT VARIANT 1030 1030 MISSING.
FT VARIANT 1094 1094 /FTID=VAR_009893.
FT VARIANT 1094 1094 L -> V.
FT VARIANT 1094 1094 /FTID=VAR_009894.
SQ SEQUENCE 1189 AA; 133609 MW; 60303084AAACD52 CRC64;

Query Match 1.1%; Score 13; DB 1; Length 1189;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181
DB 166 VIVLDGNSIYPM 178

RESULT 6
CSTA_HELPJ STANDARD; PRT; 687 AA.
ID CSTA_HELPJ STANDARD; PRT; 687 AA.
AC Q9ZK47;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbon starvation protein A homolog.
GN CSTA OR JHP1095.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deGange B.L., Carmel G.,
RA Tumano P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.";
```

```

RL Nature 397:176-180(1999).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -|- SIMILARITY: BELONGS TO THE CSTA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, AE001537; AAD06677.1; -.
CC DR InterPro: IPR003706; CSFA.
CC DR Pfam: PF02554; CSFA; 1.
KW Transmembrane, Inner membrane, Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 190 210 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 440 460 POTENTIAL.
FT TRANSMEM 462 482 POTENTIAL.
FT TRANSMEM 513 533 POTENTIAL.
FT TRANSMEM 546 566 POTENTIAL.
FT TRANSMEM 574 594 POTENTIAL.
FT TRANSMEM 642 662 POTENTIAL.
SQ SEQUENCE 687 AA; 74151 MW; A8BA383DEDE8807 CRC64;

Query Match 0.9%; Score 11; DB 1; Length 687;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 IWLILGSVLGG 1133
DB 118 IWLILGSVLGG 128

RESULT 7
CSTA_HELPJ STANDARD; PRT; 687 AA.
ID CSTA_HELPJ STANDARD; PRT; 687 AA.
AC P56190;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbon starvation protein A homolog.
GN CSTA OR HPI168.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirnes E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weiman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori.";
```

RL Nature 388:539-547(1997).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (Potential).  
CC -1- SIMILARITY: BELONGS TO THE CSTA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB000622; AAD08212.1; -.  
CC TIGR; HPI168; -.  
DR InterPro: IPR003706; Csta.  
DR Pfam: PF02554; Csta; 1.  
KW Transmembrane; Inner membrane; Complete proteome.  
FT TRANSMEM 6 26 POTENTIAL.  
FT TRANSMEM 29 49 POTENTIAL.  
FT TRANSMEM 87 107 POTENTIAL.  
FT TRANSMEM 117 137 POTENTIAL.  
FT TRANSMEM 162 182 POTENTIAL.  
FT TRANSMEM 190 POTENTIAL.  
FT TRANSMEM 221 241 POTENTIAL.  
FT TRANSMEM 256 276 POTENTIAL.  
FT TRANSMEM 285 305 POTENTIAL.  
FT TRANSMEM 325 345 POTENTIAL.  
FT TRANSMEM 373 393 POTENTIAL.  
FT TRANSMEM 440 460 POTENTIAL.  
FT TRANSMEM 462 482 POTENTIAL.  
FT TRANSMEM 513 533 POTENTIAL.  
FT TRANSMEM 546 566 POTENTIAL.  
FT TRANSMEM 574 594 POTENTIAL.  
FT TRANSMEM 642 662 POTENTIAL.  
SQ SEQUENCE 687 AA; 74204 MW; C525D97653E69C3D CRC64;  
  
Query Match 0.9%; Score 11; DB 1; Length 687;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1123 LMIIGSVLGG 1133  
Db 118 LMIIGSVLGG 128  
  
RESULT 8  
ID ITA2\_BOVIN STANDARD; PRT; 1170 AA.  
AC P53710;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)  
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).  
GN ITGA2.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
OC Bovidae; Bovinae; Bos.  
OK NCBI\_TaxID=9913;  
RX MEDLINE=94193647; PubMed=7511592;  
RP SEQUENCE FROM N.A.  
RA Kamata T., Puzon W., Takada Y.;  
RT Identification of putative ligand binding sites within I domain of  
RT integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";  
RL J. Biol. Chem. 269:9659-9663(1994).  
CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ.  
CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT  
CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN  
CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER  
CC CELLS TO COLLAGEN, MODULATION OF COLLAGEN AND COLLAGENASE GENE

CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
CC EXTRACELLULAR MATRIX.  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
CC ASSOCIATES WITH BETA-1.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC WITH I-DOMAIN. THE INTEGRIN I-DOMAIN (INSERT) IS A VMPA DOMAIN. INTEGRINS  
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 VMPA DOMAIN.  
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L25886; AAB59255.1; -.  
CC HSRP; P17301; IAOX.  
DR InterPro: IPR000413; Integrin\_alpha.  
DR InterPro: IPR002035; VMP\_A.  
DR Pfam; PF00092; vva; 1.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR SMART; SMO0191; FG-GAP; 5.  
DR SMART; SMO0327; VMA; 1.  
DR SMART; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS50234; VMPA; 1.  
KW Integrin; Cell adhesion; Receptor; Polymorphism; Calcium; Magnesium.  
FT SIGNAL 1 1  
FT CHAIN 19 18 POTENTIAL.  
FT DOMAIN 19 1170 INTEGRIN ALPHA-2.  
FT TRANSMEM 1122 1143 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 1144 1170 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 34 92 FG-GAP 1.  
FT REPEAT ? 7 FG-GAP 2.  
FT DOMAIN 177 367 VMPA.  
FT REPEAT ? 367 FG-GAP 3.  
FT REPEAT 423 475 FG-GAP 4.  
FT REPEAT 477 538 FG-GAP 5.  
FT REPEAT 540 599 FG-GAP 6.  
FT REPEAT 604 656 FG-GAP 7.  
FT CA\_BIND 488 496 POTENTIAL.  
FT CA\_BIND 552 560 POTENTIAL.  
FT CA\_BIND 616 624 POTENTIAL.  
FT SITE 472 474 GEFPR MOTIF.  
FT SITE 1146 1150 CELL ATTACHMENT SITE (POTENTIAL).  
FT DISULFID 72 81 BY SIMILARITY.  
FT DISULFID 669 726 BY SIMILARITY.  
FT DISULFID 778 784 BY SIMILARITY.  
FT DISULFID 854 865 BY SIMILARITY.  
FT DISULFID 1008 1039 BY SIMILARITY.  
FT DISULFID 1044 1049 BY SIMILARITY.  
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1063 1063 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 580 580 G -> V.  
FT VARIANT 588 588 R -> K.  
FT VARIANT 725 725 R -> S.  
SQ SEQUENCE 1170 AA; 128929 MW; EECEP1C5F2448FB1 CRC64;

Query Match 0.9%; Score 11; DB 1; Length 1170;  
 Best Local Similarity 100.0%; Pred. No. 0.029;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHG 282  
 DB 268 VVVTGESHG 278

RESULT 9  
 ITA2\_MOUSE STANDARD; PRT; 1178 AA.  
 AC 062469; 062163;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GPIa)  
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).  
 GN ITGA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxId=10090;  
 RX MEDLINE=94355691; PubMed=8081889;  
 RA Edelman J.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,  
 RA Damjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;  
 RA "The mouse VLA-2 homologue supports collagen and laminin adhesion but  
 RT not virus binding.";  
 RL Cell Adhes. Commun. 2:131-143(1994).  
 RP [2]  
 RP SEQUENCE OF 450-1178 FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=94355691; PubMed=7521231;  
 RA Wu J.E., Santoro S.A.;  
 RA "Complex patterns of expression suggest extensive roles for the alpha  
 RT 2 beta 1 integrin in murine development.";  
 RL Dev. Dyn. 199:392-314(1994).  
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A COLLAGEN RECEPTOR, BEING  
 CC RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER CELLS TO  
 CC COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE EXPRESSION,  
 CC FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
 CC EXTRACELLULAR MATRIX. IT IS ALSO A RECEPTOR FOR LAMININS, COLLAGEN  
 CC C-PROPEPTIDES AND E-CADHERIN. MICE HOMOLOGOUS FOR A NULL MUTATION  
 CC IN THE ALPHA-2 DIE VERY EARLY IN EMBRYOGENESIS.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z29987; CA82877.1; -;  
 DR EMBL; X75427; CA53178.1; -;  
 DR HSSP; P17301; IAOX.  
 DR MGD; MGI:96600; Itga2.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF000692; vwa; 1.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF01839; FG-GAP; 5.

DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA\_1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS50334; VWFA; 1.  
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Platelet; Signal; Repeat; Calcium; Magnesium.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1178  
 FT DOMAIN 27 1129  
 FT TRANSMEM 1130 1151  
 FT DOMAIN 1152 1178  
 FT REPEAT 42 100  
 FT REPEAT ? ?  
 FT DOMAIN 185 375  
 FT REPEAT ? ?  
 FT REPEAT 431 483  
 FT REPEAT 485 546  
 FT REPEAT 548 607  
 FT REPEAT 612 664  
 FT CA\_BIND 496 504  
 FT CA\_BIND 560 568  
 FT CA\_BIND 624 632  
 FT SITE 480 482  
 FT SITE 1154 1158  
 FT DISULFID 80 89  
 FT DISULFID 677 734  
 FT DISULFID 786 792  
 FT DISULFID 862 873  
 FT DISULFID 1016 1047  
 FT DISULFID 1052 1057  
 FT CARBOHYD 102 102  
 FT CARBOHYD 109 109  
 FT CARBOHYD 429 429  
 FT CARBOHYD 457 457  
 FT CARBOHYD 472 472  
 FT CARBOHYD 696 696  
 FT CARBOHYD 1054 1054  
 FT CARBOHYD 1071 1071  
 FT CARBOHYD 1078 1078  
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;  
 Query Match 0.9%; Score 11; DB 1; Length 1178;  
 Best Local Similarity 100.0%; Pred. No. 0.029;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 272 VVVTGESHG 282  
 DB 276 VVVTGESHG 286  
 RESULT 10  
 ITA2\_HUMAN STANDARD; PRT; 1181 AA.  
 AC P17301;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GPIa)  
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).  
 GN ITGA2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxId=9606;  
 RX MEDLINE=89308879; PubMed=2545729;  
 RA Takeda Y., Hemler M.E.;  
 RA "The primary structure of the VLA-2/collagen receptor alpha 2 subunit  
 RT (platelet GPIa): homology to other integrins and the presence of a  
 RT possible collagen-binding domain.";

RL J. Cell Biol. 109:397-407(1989).  
 RN (2)  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.  
 RX MEDLINE=98019223; PubMed=9353312;  
 RA Emsley J., King S.L., Bergelson J.M., Liddington R.C.;  
 RT "Crystal structure of the I domain from integrin alpha2beta1";  
 RL J. Biol. Chem. 272:28512-28517(1997).  
 RN (3)  
 RP VARIANT HPA-5 (BR).  
 RX MEDLINE=94043762; PubMed=7901236;  
 RA Santoso S., Kalb R., Walke M., Kiefel V., Mueller-Eckhardt C.,  
 RT Newman P.J.;  
 RT "The human platelet alpha2beta1 integrin is associated with a  
 single amino acid polymorphism on glycoprotein Ia (integrin subunit  
 alpha 2).";  
 RL J. Clin. Invest. 92:2427-2432(1993).  
 RN (4)  
 RP VARIANT GLU-534.  
 RX MEDLINE=2026009; PubMed=10744142; Haberbosch W., Santoso S.;  
 RA Kroll H., Gardemann A., Fechter A.,  
 RT "The impact of the glycoprotein Ia collagen receptor subunit A1648G  
 gene polymorphism on coronary artery disease and acute myocardial  
 infarction.";  
 RL Thromb. Haemost. 83:392-396(2000).  
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,  
 COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT  
 RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN  
 COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER  
 CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE  
 EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
 EXTRACELLULAR MATRIX.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
 ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INTEGR) IS A VFMA DOMAIN. INTEGRINS  
 WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- POLYMORPHISM: POSITION 534 IS ASSOCIATED WITH PLATELET-SPECIFIC  
 ALLOANTIGEN HPA-5 (BR). HPA-5A/BR(A) HAS LYS-534 AND HPA-5B/BR(B)  
 HAS GLU-534. HPA-5B IS INVOLVED IN NEONATAL ALLOIMMUNE  
 THROMBOCYTOPENIA (NMIT OR NATP). THE K534E POLYMORPHISM MAY PLAY A  
 ROLE IN CORONARY ARTERY DISEASE (CAD).  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -1- DATABASE: NAME=PRO; NOTE=CD guide CD49b entry;  
 WWW="http://www.ncbi.nlm.nih.gov/ncbi/cd/cd49b.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@sib-sib.ch).  
 CC -----  
 CC EMBL: X17033; CAA34894.1; -  
 CC PIR: A33998; A33998.  
 CC PDB: 1A0X; 25-NOV-98.  
 CC Genew; HGNC:6137; ITGA2.  
 DR MIM; 192974; -  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VFMA\_A.  
 DR Pfam; PF00092; vwa; 1.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR PRINTS; PR00453; VFMADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; vwa; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS50234; VFMA; 1.  
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 Platelet; Signal; Repeat; Calcium; Magnesium; Polymorphism;  
 3D-structure.

FT	SIGNAL	1	29	
FT	CHAIN	30	1181	INTEGRIN ALPHA-2.
FT	DOMAIN	30	1132	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1133	1154	POTENTIAL.
FT	DOMAIN	1155	1181	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	45	103	FG-GAP 1.
FT	REPEAT	?	?	FG-GAP 2.
FT	DOMAIN	188	378	VFMA.
FT	REPEAT	378	433	FG-GAP 3.
FT	REPEAT	434	486	FG-GAP 4.
FT	REPEAT	488	549	FG-GAP 5.
FT	REPEAT	551	610	FG-GAP 6.
FT	REPEAT	615	667	FG-GAP 7.
FT	CA_BIND	499	507	POTENTIAL.
FT	CA_BIND	563	571	POTENTIAL.
FT	CA_BIND	627	635	POTENTIAL.
FT	SITE	1157	1161	GFPR MOTIF.
FT	DISULFID	83	92	BY SIMILARITY.
FT	DISULFID	680	737	BY SIMILARITY.
FT	DISULFID	789	795	BY SIMILARITY.
FT	DISULFID	865	876	BY SIMILARITY.
FT	DISULFID	1019	1050	BY SIMILARITY.
FT	DISULFID	1055	1060	BY SIMILARITY.
FT	CARBOHYD	105	105	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	343	343	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	432	432	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	475	475	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	699	699	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1057	1057	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1074	1074	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1081	1081	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	534	534	K -> E (IN ALLOANTIGEN HPA-5B).
FT	SEQUENCE	1181 AA;	129295 MW;	7E1B7ED968A94070 CRC64;
FT	Query Match	0.9%;	Score 11;	DB 1; Length 1181;
FT	Best Local Similarity	100.0%;	Pred. No. 0.029;	
FT	Matches 11;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	272	VVVTGDSHSDG	282	
Db	279	VVVTGDSHSDG	289	
RESULT 11				
ID	ITAE_HUMAN	STANDARD;	PRT;	1179 AA.
AC	P38570; Q9NZU9;			
DT	01-OCT-1994 (Rel. 30. Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Integrin alpha-E precursor (Mucosal lymphocyte-1 antigen) (HML-1 antigen) (CD103 antigen) (Integrin alpha-IEH).			
GN	ITGA2			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 19-38 AND 179-188.			
RC	TISSUE=Lymphocytes, and Leukemia;			
RX	MEDLINE=94164962; PubMed=8119477;			
RA	Shaw S.K., Cepek K.L., Murphy E.A., Russell G.J., Brenner M.B.,			
RT	"Molecular cloning of the human mucosal lymphocyte integrin alpha E			
RT	subunit. Unusual structure and restricted RNA distribution.";			
RL	J. Biol. Chem. 269:6016-6025(1994).			
RN	[2]			
RP	REVISIONS TO 88-114.			
RA	Parker C.M.;			
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.			

RN	[3]
RP	SEQUENCE OF S3-1179 FROM N.A.
RC	TISSUE=Fetal kidney;
RX	MEDLINE=20138496; PubMed=10673275; Tschuchnig J.W., Ankster Y., Dietrich N.L., Maduro V.V., McDowell G., RA Shocler-Bard V., Bouffard G.G., Beckstrom-Sternberg S.M., Gahl W.A., RA Green E.D.;
RT	"The genomic region encompassing the nephropathic cystinosis gene (CTNS) : complete sequencing of a 200-kb segment and discovery of a novel gene within the common cytostasis-causing deletion."
RM	Genome Res. 10:165-173(2000).
RL	(4)
RP	MUTAGENESIS OF ASP-109 AND PHE-316. RX MEDLINE=20400502; PubMed=10837471; RA Higgins J.M.H., Cernadas M., Tan K., Irie A., Wang J.-H., Takada Y., RA Brenner M.B.;
RT	"the role of alpha and beta chains in ligand recognition by beta 7 integrins." ;
RL	J. Biol. Chem. 275:25652-25664(2000). CC - FUNCTION: INTEGRIN ALPHA-E/BETA-7 IS A RECEPTOR FOR E-CADHERIN. IT MEDIATES ADHESION OF INTRA-EPITHELIAL T-LYMPHOCYTES TO EPITHELIAL CELL MONOLAYERS.
CC	- SUBUNIT: HETERO DIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAINS LINKED BY A DISULFIDE BOND. ALPHA-E ASSOCIATES WITH BETA-7.
CC	- SUBCELLULAR LOCATION: Type I membrane protein.
CC	- TISSUE SPECIFICITY: EXPRESSED ON A SUBCLASS OF T-LYMPHO CYTES KNOWN AS INTRA-EPITHELIAL LYMPHO CYTES WHICH ARE LOCATED BETWEEN MU COSAL EPITHELIAL CELLS.
CC	- DOMAIN: THE INTERGRIN I-DOMAIN (INSERT) IS A WFPA DOMAIN. INTEGRINS WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC	- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC	- SIMILARITY: CONTRAINS 1 WFPA DOMAIN.
CC	- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC	- DATABASE: NAME=PROW; NOTE=CD guide CD103 entry: <a href="http://www.ncbi.nlm.nih.gov/prow/cd/cdt03.htm">WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cdt03.htm"</a> .
CC	-- -- -- -- --
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its con- tent or use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by end user and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/">http://www.isb-sib.ch/</a> ) or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ). -----
DR	EMBL, L25851; AAB59359.2; -. DR EMBL, AF168787; AAP43107.1; -. DR HSSP; HGNC:6147; ITGAE. DR Genev; HGNC:6147; ITGAE. DR MIR_604682; .
DR	InterPro; IPRO00413; Integrin_alpha. DR InterPro; IPRO02035; VWFA_A. DR Pfam; PF00092; vwa; 1. DR Pfam; PF00357; integrin_A; 1. DR Pfam; PF01839; FG-GAP; 4. DR PRINTS; PR01185; INTEGRINA. DR PRINTS; PR00453; VFPRDOMAIN. DR SMART; SM00191; Int_alpha; 4. DR SMART; SM00327; VWFA; 1. DR PROSITE; PS00242; INTEGRIN_ALPHA; 1. DR PROSITE; PSS0234; VWFA; 1. KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane; KW Signal; Repeat; Polymorphism; Magnesium; Calcium. FT SIGNAL 1 .. 18 FT CHAIN 19 .. 1179 INTEGRIN ALPHA-E. FT CHAIN 19 .. 177 INTEGRIN ALPHA-E LIGHT CHAIN. FT CHAIN 179 .. 1177 INTEGRIN ALPHA-E HEAVY CHAIN. FT CHAIN 19 .. 1124 EXTRACELLULAR (POTENTIAL). FT TRANSMEM 1125 .. 1147 POTENTIAL. FT DOMAIN 1148 .. 1179 CYTOPLASMIC (POTENTIAL). FT DOMAIN 181 .. 198 GUU-RICH (ACIDIC). FT REPEAT ? .. ? FG-GAP 1. FT REPEAT ? .. ? FG-GAP 2.

Query Match	Best Local Similarity	0.9%: Score 10; DB 1; Length 1179;
Matches 10; Conservative	100.0%; Pred. No. 0.3;	
	0; Mismatches	0; Indels
		0; Gaps
		0
QY 483 GSYFGSELCP 492		
DB 511 GSYFGSELCP 520		
RESULT 12		
ITAL_MOUSE	STANDARD;	PRT; 1163 AA.
AC P24063;		
DT 01-MAR-1992 (Rel. 21, Created)		
DT 01-MAR-1992 (Rel. 21, Last sequence update)		
DT 15-JUN-2002 (Rel. 41, Last annotation update)		
DE Integrin alpha-L precursor (Leukocyte adhesion glycoprotein LFA-1 alpha chain) (Leukocyte function associated molecule 1, alpha chain) (CD11a).		
DE ITGAL OR LFA-1.		
GN Mus musculus (Mouse).		
OS Mus musculus (Mouse).		
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX NCBI_TaxID=10090;		
OX [1]		
RN SEQUENCE FROM N.A.		
RX MEDLINE=91268576; PubMed=2051027;		
RA Kaufmann Y., Tseng E., Springer T.A.;		
RT "Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit and its expression in COS cells.";		
RL J. Immunol. 147:369-374 (1991).		
RN [12]		

RX	SEQUENCE OF 24-42.
RA	MEDLINE=85188276; PubMed=3887182;
RT	Springer T.A., Teplow D.B., Dreyer W.J.;
RL	"Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion glycoproteins and unexpected relation to leukocyte interferon"; Nature 314:540-542(1985).
CC	- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2, ICAM3 AND ICAM4. IS INVOLVED IN A VARIETY OF IMMUNE PHENOMENA INCLUDING LEUKOCYTE-ENDOTHELIAL CELL INTERACTION, CYTOTOXIC T-CELL MEDIATED KILLING, AND ANTIBODY DEPENDENT KILLING BY GRANULOCYTES AND MONOCYTES. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-L SUBUNIT GENE DEMONSTRATE IMPAIRED TUMOR REJECTION AND IMPAIRED LEUKOCYTES RECRUITMENT.
CC	- SUBUNIT: HETRODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-L ASSOCIATES WITH BETA-2.
CC	- SUBCELLULAR LOCATION: Type I membrane protein.
CC	- TISSUE SPECIFICITY: LEUKOCYTES.
CC	- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC	- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC	- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC	- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC	----- THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-stb.ch/announce/or send an email to license@isb-stb.ch). -----
DR	EMBL; M60778; AAA39426.1; -.
DR	HSSP; P20701; ILPA.
DR	MCD; MG1:96606; Itgal.
DR	InterPro: IPRO00413; Integrin_alpha.
DR	InterPro: IPR002035; VWF_A.
DR	Pfam; PF00092; vwa; 1.
DR	Pfam; PF00357; Integrin_A; 1.
DR	Pfam; PF01839; FG-GAP 5.
DR	PRINTS; PR01185; INTEGRINA.
DR	PRINTS; PR00453; VWFADOMAIN.
DR	SMART; SM00191; Int_alpha; 5.
DR	SMART; SM00327; VWFA; 1.
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR	PROSITE; PS50234; VWFA; 1.
KW	Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane; Signal; Magnesium; Calcium;
KW	Repeat.
FT	SIGNAL. 1 23
FT	CHAIN 24 1163
FT	DOMAIN 24 1084 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 1085 1108 POTENTIAL.
FT	DOMAIN 1109 1163 CYTOPLASMIC (POTENTIAL).
FT	REPEAT 39 88 FG-GAP 1.
FT	REPEAT ? ? FG-GAP 2.
FT	DOMAIN 148 334 ? ? VWFA.
FT	REPEAT ? ? FG-GAP 3.
FT	REPEAT 399 454 FG-GAP 4.
FT	REPEAT 455 514 FG-GAP 5.
FT	REPEAT 516 573 FG-GAP 6.
FT	REPEAT 576 628 FG-GAP 7.
FT	CA_BIND 466 474 POTENTIAL.
FT	CA_BIND 528 536 POTENTIAL.
FT	CA_BIND 588 596 POTENTIAL.
FT	CITE SITE 1111 1115 GPCR MOTIF.
FT	DISULFID 70 77 BY SIMILARITY.
FT	DISULFID 108 126 BY SIMILARITY.
FT	DISULFID 147 199 BY SIMILARITY.
FT	DISULFID 651 705 BY SIMILARITY.
FT	DISULFID 767 773 BY SIMILARITY.
FT	DISULFID 840 856 BY SIMILARITY.
FT	DISULFID 993 1009 BY SIMILARITY.
FT	DISULFID 1017 1048 BY SIMILARITY.

FT	CARBOHYD	86	86	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	185	185	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	270	270	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	444	444	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	668	668	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	624	624	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	728	728	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	776	776	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	857	857	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	880	880	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	890	890	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	899	899	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	927	927	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
FT	CARBOHYD	1056	1056	N-LINKED	(GLCNAC. . .)	(POTENTIAL)		
SQ	SEQUENCE	1163 AA;	128343 MW;	A7A3078489E8232F	CRC64;			
Query Match								
Best Local Similarity		0.8%;	Score 9;	DB 1;	Length 1163;			
Matches		9;	Conservative	0;	Mismatches	0;		
					Indels	0;		
					Gaps			
Oy	566 DVAVGAPLE	574						
Db	536 DVAVGAPLE	544						
RESULT 13								
ID	ITAL_HUMAN	STANDARD;	PRT;	1170 AA.				
AC	P20701; O43746;							
DT	01-FEB-1991 (Rel. 17, Created)							
DT	01-NOV-1995 (Rel. 32, Last sequence update)							
DT	15-JUN-2002 (Rel. 41, Last annotation date)							
DE	Integrin alpha-L precursor (leukocyte adhesion glycoprotein LFA-1							
DE	alpha chain) (leukocyte function associated molecule 1, alpha chain)							
DE	(CD11a).							
GN	ITGAL OR CD11A.							
OS	Homo sapiens (Human).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
OX	NCBI_TaxID=9606;							
RN	SEQUENCE FROM N.A. (ISOPFORM 1), AND PARTIAL SEQUENCE.							
RP	MEDLINE=69139587; PubMed=2537322;							
RX	"Pamson R.S., Corbi A.L., Berman L., Springer T.;							
RA	"Primary structure of the leukocyte function-associated molecule-1							
RT	alpha subunit: an integrin with an embedded domain defining a protein							
RT	superfamily.";							
RL	J. Cell Biol. 108:703-712(1989).							
RN								
RP	SEQUENCE FROM N.A. (ISOPFORM 2).							
RX	MEDLINE=99425270; PubMed=10493829;							
RA	Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,							
RA	Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,							
RA	DeLalates Mays A., Cao Y., Xu R.X., Kang H.-D., Mitchell S.,							
RA	Eschler E.E., Harris P.C., Venter J.C., Adams M.D.;							
RA	"Genome duplications and other features in 12 Mb of DNA sequence from							
RT	human chromosome 16p and 16q.";							
RL	Genomics 60:295-308(1999).							
RN								
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 153-335, AND REVISION TO 214							
RX	MEDLINE=96036067; PubMed=7479767;							
RA	Ou A., Leahy D.J.;							
RT								







RX MEDLINE=94043127; PubMed=8226859;  
 RA Wenger R.H., Rochelle J.M., Seldin M.F., Koehler G., Nielsen P.U.;  
 RT "The heat stable antigen (mouse CD24) gene is differentially  
 RT regulated but has a housekeeping promoter.";  
 RL J. Biol. Chem. 268:23345-23352(1993).  
 RN [5]  
 RP SEQUENCE OF 27-53.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=92412120; PubMed=1530634;  
 RA Hatakeyama Y., Nakano A., Ohnishi H., Hamada F., Saheki S.,  
 RA Takeuchi N.;  
 RT "Purification of the murine heat-stable antigen from erythrocytes";  
 RL Biochem. Biophys. Res. Commun. 187:773-777(1992).  
 CC -1- FUNCTION: MAY HAVE A SPECIFIC ROLE TO PLAY IN EARLY THYMOCYTE  
 CC DEVELOPMENT.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: IN LYMPHOID, MYELOID, AND ERYTHROID CELLS.  
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, MS8661; AAA39481.1; -  
 DR EMBL, X56469; CA39841.1; -  
 DR EMBL, X72910; CA51415.1; -  
 DR EMBL, X53825; CA37822.1; -  
 DR PIR, S43709; S43709.  
 DR PIR, A43537; A43537.  
 DR PIR, S32240; S32240.  
 DR PIR, S15784; S15784.  
 DR PIR, S33129; S33129.  
 DR PIR, S19111; S19111.  
 DR MGI:88323; Cd24a.  
 KW Glycoprotein; GPI-anchor; Membrane; Signal; Antigen.  
 FT SIGNAL 1 26  
 FT CHAIN 27 53  
 FT PROPEP 54 76  
 FT CARBOHYD 27 27  
 FT CARBOHYD 30 30  
 FT CARBOHYD 39 39  
 FT CARBOHYD 48 48  
 FT LIPID 53 53  
 SQ SEQUENCE 76 AA; 7797 MW; 6853F12183625EB CRC64;  
 Query Match 0.7%; Score 8; DB 1; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1133 GILLIAL 1140  
 Db 12 GILLIAL 19  
 RESULT 15  
 ID CD24 RAT STANDARD; PRT; 76 AA.  
 AC 007450;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Signal transducer CD24 precursor (Heat stable antigen) (HSA)  
 DE (Nectadrin).  
 DE CD24A.  
 GN Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 OK

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WiStar; TISSUE=Embryonic brain;  
 RX MEDLINE=94122434; PubMed=8292828;  
 RA Shiraesawa T., Akashi T., Sakamoto K., Takahashi H., Maruyama N.,  
 RA Hirokawa K.;  
 RT "Gene expression of CD24 core peptide molecule in developing brain  
 RT and developing non-neural tissues.";  
 RL Dev. Dyn. 198:1-13(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fischer;  
 RX MEDLINE=97157759; PubMed=9004038;  
 RA Magnaldo T.A., Barandon Y.;  
 RT "CD24 (heat stable antigen, nectadrin), a novel keratinocyte  
 RT differentiation marker, is preferentially expressed in areas of the  
 RT hair follicle containing the colony-forming cells";  
 RL J. Cell Sci. 109:3035-3045(1996).  
 CC -1- FUNCTION: MAY HAVE A PIVOTAL ROLE IN CELL DIFFERENTIATION. THE  
 CC TRIGGERING MECHANISM OF SIGNAL TRANSDUCTION MAY BE DUE TO THE  
 CC INTERACTIONS OF DIFFERENTIATING CELLS WITH THE MATRIX SUBSTRATE  
 CC VIA THE CARBOHYDRATE STRUCTURE OF THE MOLECULE. IN THIS WAY, THE  
 CC SIGNAL TRANSDUCER CAN PLAY VERY DIFFERENT ROLES IN DIFFERENT CELL  
 CC TYPES AS A DIRECT CONSEQUENCE OF ITS GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE CENTRAL NERVOUS SYSTEM, IN  
 CC POSTMITOTIC CELLS OF SPINAL CORD, HINDBRAIN, MIDBRAIN AND  
 CC FOREBRAIN. EXPRESSED IN EPITHELIUM DURING THE DEVELOPMENT OF NON-  
 CC NEURAL TISSUES. EXPRESSED IN TOOTH DEVELOPMENT, SPECIFICALLY IN  
 CC MESENCHYMAL CELLS DIFFERENTIATING INTO ODONTOBLAST IN DENTAL  
 CC PAPILLA, AS WELL AS IN THE DEVELOPING EYE AND HAIR FOLLICLE.  
 CC -1- DEVELOPMENTAL STAGE: DETECTED IN PRIMITIVE ECODERM, MESODERM AND  
 CC VENTRAL ENDODERM, DOWN-REGULATED WHEN ORGANOGENESIS IS COMPLETED.  
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY). THE CARBOHYDRATE  
 CC STRUCTURE MAY BE REGULATED IN A TISSUE SPECIFIC AND DEVELOPMENTAL  
 CC STAGE SPECIFIC MANNER.  
 CC -1- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, Z11663; CA77731.1; -  
 DR EMBL, U49062; AAA91470.1; -  
 KW Glycoprotein; GPI-anchor; Membrane; Signal; Differentiation.  
 FT SIGNAL 1 26  
 FT CHAIN 27 56  
 FT PROPEP 57 76  
 FT CARBOHYD 27 27  
 FT CARBOHYD 37 37  
 FT CARBOHYD 48 48  
 FT LIPID 56 56  
 SQ SEQUENCE 76 AA; 7862 MW; 42846E70EC39D958 CRC64;  
 Query Match 0.7%; Score 8; DB 1; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1133 GILLIAL 1140  
 Db 12 GILLIAL 19  
 RESULT 16  
 ID CD24 HUMAN STANDARD; PRT; 80 AA.  
 AC P25063; Q16257;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Signal transducer CD24 precursor.  
 GN CD24 OR CD24.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91332458; PubMed=1831224;  
 RA Kay R., Roeten P.M., Humphries R.K.;  
 RT "CD24, a signal transducer modulating B cell activation responses, is  
 RT a very short peptide with a glycosyl phosphatidylinositol membrane  
 RT anchor";  
 RL J. Immunol. 147:1412-1416(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93007871; PubMed=1327504;  
 RA Jackson D., Walbel R., Weber E., Bell J., Stachel R.A.;  
 RT "CD24, a signal-transducing molecule expressed on human B cells, is a  
 RT major surface antigen on small cell lung carcinomas";  
 RL Cancer Res. 52:5264-5270(1992).  
 RN [3]  
 RP SEQUENCE OF 1-76 FROM N.A.  
 RX MEDLINE=95048364; PubMed=7959762;  
 RA Hough M.R., Roeten P.M., Sexton T.L., Kay R., Humphries R.K.;  
 RT "Mapping of CD24 and homologous sequences to multiple chromosomal  
 RT loci";  
 RL Genomics 22:154-161(1994).  
 CC -1- FUNCTION: MODULATES B-CELL ACTIVATION RESPONSES. SIGNALING COULD  
 CC BE TRIGGERED BY THE BINDING OF A LECTIN-LIKE LIGAND TO THE CD24  
 CC CARBOHYDRATES, AND TRANSDUCED BY THE RELEASE OF SECOND MESSAGES  
 CC DERIVED FROM THE GPI-ANCHOR. PROMOTES AG-DEPENDENT PROLIFERATION  
 CC OF B-CELLS. AND PREVENTS THEIR TERMINAL DIFFERENTIATION INTO  
 CC ANTIBODY-FORMING CELLS.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: B-CELLS.  
 CC -1- INDICATION: EXPRESSION IS LOST WHEN PRIMARY B-CELLS ARE INDUCED TO  
 CC DIFFERENTIATE IN ANTIBODY-FORMING CELLS.  
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.  
 CC -1- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.  
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD24 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd24.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: M58664; AAA35665.1; -;  
 CC EMBL: X69397; CA449195.1; -;  
 CC EMBL: L33930; AAB58807.1; -;  
 CC EMBL: S75311; AAD14170.1; ALT\_INIT.  
 CC DR Genew; HGNC:1645; CD24.  
 CC MIM: 600074; -;  
 CC DR KW Glycoprotein; GPI-anchor; Membrane; Signal.  
 CC FT SIGNAL 1 26 POTENTIAL.  
 CC FT CHAIN 27 59 SIGNAL TRANSDUCER CD24.  
 CC FT PROPEP 60 80 REMOVED IN NATURE FORM (POTENTIAL).  
 CC FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT LIPID 59 59 GPI-ANCHOR (POTENTIAL).  
 CC FT CONFLICT 12 12 G -> W (IN REF. 3).  
 CC FT CONFLICT 44 44 S -> T (IN REF. 3).  
 CC FT CONFLICT 57 57 A -> V (IN REF. 2).  
 CC FT CONFLICT 57 57 A -> V (IN REF. 2).  
 CC FT SEQUENCE 80 AA; 8083 MW; DB1988B6808F833F CRC64;  
 SO

OY 1133 GULLBALL 1140  
 DB 12 GULLBALL 19  
 RESULT 17  
 ID PRVB MERBI STANDARD; PRT; 108 AA.  
 AC P56503;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Parvalbumin beta.  
 OS Merluccius bilinearis (Silver hake).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Paracanthopterygii; Gadiformes; Merlucciidae;  
 OC Merluccius.  
 OX NCBI\_TaxID=79698;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=98046752; PubMed=9385642;  
 RA Revett S.P., King G., Shabanowitz J., Hunt D.F., Hartman T.M.,  
 RA Nelson D.J.;  
 RT "Characterization of a helix-loop-helix (EF hand) motif of silver  
 RT hake parvalbumin isoform B";  
 RL Protein Sci. 6:2397-2408(1997).  
 CC -1- FUNCTION: IN MUSCLE, THE CALCIUM-BINDING PROTEIN PARVALBUMIN IS  
 CC THOUGHT TO BE INVOLVED IN MUSCLE RELAXATION.  
 CC -1- MASS SPECTROMETRY: MW=11357; MW ERR=3.5; METHOD=Electrospray.  
 CC -1- MISCELLANEOUS: THIS PARVALBUMIN HAS AN ISOELECTRIC POINT OF 4.2.  
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS, BELONGS TO  
 CC THE PARVALBUMINS SUBFAMILY.  
 CC HSP: P02621; 1475.  
 CC DR InterPro; IPR002048; EF-hand.  
 CC DR Pfam; PF00036; ehand; 2.  
 CC DR ProDom; PD000012; EF-hand; 1.  
 CC DR PROSITE; PS00018; EF-HAND; 2.  
 CC KW Calcium-binding; Muscle protein; Repeat; Acetylation.  
 CC FT MOD RES 1 1 ACETYLATION (PROBABLE).  
 CC FT CA BIND 51 62 EF-HAND 1 (BY SIMILARITY).  
 CC FT CA BIND 90 101 EF-HAND 2 (BY SIMILARITY).  
 CC FT SEQUENCE 108 AA; 11317 MW; 08788BFC54002906 CRC64;  
 SO

Query Match 0.7%; Score 8; DB 1; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 4.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 287 AALKACEA 294  
 DB 13 AALKACEA 20  
 RESULT 18  
 ID 1122 HUMAN STANDARD; PRT; 179 AA.  
 AC Q9GZX6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Interleukin-22 precursor (IL-22) (IL-10-related T-cell-derived  
 DE inducible factor) (IL-TIF).  
 GN IL22 OR ILTIF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20420346; PubMed=10954742;  
 RA Dumoutier L., Van Roost E., Colau D., Renaud J.-C.,  
 RA "Human interleukin-10-related T cell-derived inducible factor";

RT molecular cloning and functional characterization as an hepatocyte-stimulating factor." ;  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:10144-10149(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21069354; PubMed=1197690;  
 RA Dumoutier L., Van Roost E., Coliau D., Ameys G., Michaux L., Renaud J.-C.;  
 RT "IL-TIF/IL-22: genomic organization and mapping of the human and mouse genes";  
 RL Genes Immun. 1:488-494(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20469498; PubMed=10875937;  
 RA Xie M.-H., Aggarwal S., Ho W.-H., Foster J., Zhang Z., Stinson J., Wood W. I., Goddard A. D., Gunney A. L.;  
 RT "Interleukin (IL)-22, a novel human cytokine that signals through the interferon receptor-related proteins CRF2-4 and IL-22R";  
 RL J. Biol. Chem. 275:31335-31339(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A., AND VARIANT GUY-158.  
 RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q., Nickerson D.A.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Cytokine that contributes to the inflammatory response in vivo.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE IL-10 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ277247; CAC06085.1; -  
 DR EMBL: AJ277248; CAC19409.1; -  
 DR EMBL: AF279437; AAG22064.1; -  
 DR EMBL: AF387519; AAK62468.1; -  
 DR Genew; HGNC:14900; IL22.  
 DR MIM; 605330; -  
 DR InterPro; IPR000098; Interleukin\_10.  
 DR SMART; SMO0188; IL10; 1.  
 DR PROSITE; PS00520; INTERLEUKIN\_10; 1.  
 KM Cytokine; Glycoprotein; Signal; Polymorphism.  
 FT CHAIN 1 33  
 FT SIGNAL 1 33  
 FT CARBOHYD 54 54  
 FT CARBOHYD 68 68  
 FT CARBOHYD 97 97  
 FT VARIANT 158 158  
 FT SEQUENCE 179 AA; 20011 MW; 3C35E64D60CF8767 CRC64;  
 SQ  
 Query Match 0.7%; Score 8; DB 1; Length 179;  
 Best Local Similarity 100.0%; Pred. No. 6.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GN MBP2.  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
 OC NCBI\_TaxID=10141;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Eosinophil;  
 RX MEDLINE=91224343; PubMed=2026266;  
 RA Aoki I., Shindoh Y., Nishida T., Nakai S., Hong Y.-M., Mio M., Saito T., Tasaka K.;  
 RT "Comparison of the amino acid and nucleotide sequences between human and two guinea pig major basic proteins";  
 RL FEBS Lett. 282:56-60(1991).  
 CC -1- FUNCTION: MBP MAY PLAY SOME IMPORTANT ROLES IN THE ALLERGIC REACTIONS AND INFLAMMATIONS, SINCE MBP IS CAPABLE OF RELEASING HISTAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF BRONCHIAL TUBES. ANTIPARASITIC AND ANTIBIOTIC.  
 CC -1- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC GRANULE (CRYSTALLOID CORE).  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D0817; BAA0697.1; -  
 DR PIR; S15102; S15102.  
 DR HSSP; P13727; IH8U.  
 DR InterPro; IPR002352; Emaior basic.  
 DR InterPro; IPR001304; Lectin\_C.  
 DR Pfam; PF00059; Lectin\_C; 1.  
 DR PRINTS; PR00770; EMAJORBASICP.  
 DR SMART; SMO0034; CLECT; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
 DR PROSITE; PS50041; C-TYPE LECTIN\_2; 1.  
 KM Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;  
 KM Lectin; Multigene family; Glycoprotein.  
 FT SIGNAL 1 15  
 FT PROPEP 16 115  
 FT CHAIN 116 234  
 FT DOMAIN 133 234  
 FT DISULFD 135 232  
 FT DISULFD 209 224  
 FT CARBOHYD 69 69  
 FT SEQUENCE 234 AA; 26140 MW; 7D926A942BF116F CRC64;  
 SQ  
 Query Match 0.7%; Score 8; DB 1; Length 234;  
 Best Local Similarity 100.0%; Pred. No. 8.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLALLV 1141  
 DB 21 LLLALLV 28

RESULT 20  
 PSAT YEAST  
 ID PSAT YEAST STANDARD; PRT; 254 AA.  
 AC P40303.  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Proteasome component PR6 (EC 3.4.25.1) (Macropain subunit PR6) (Proteinase YSCE subunit PR6) (Multicatalytic endopeptidase complex subunit PR6).  
 DE PR6 OR YOL038W.  
 GN Saccharomyces cerevisiae (Baker's yeast).  
 OS

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces;  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C;  
 RX MEDLINE=95001940; PubMed=7918444;  
 RA Heinemeyer W., Troendle N., Albrecht G., Wolf D.H.;  
 RT "PRE5 and PRE6, the last missing genes encoding 20S proteasome  
 subunits from yeast? Indication for a set of 14 different subunits in  
 the eukaryotic proteasome core.";  
 RL Biochemistry 33:12229-12237(1994).  
 RP SEQUENCE FROM N.A.  
 RA Ansoorge W., Bernes V., Rechmann S., Schwager C., Teodoru C., Voss H.,  
 RA Wiemann S.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Habbig B., Hattenhorst U., Hollenberg C.P., Ramezani Rad M.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 112-133.  
 RA Troendle N.;  
 RL Thesis (1991), University of Stuttgart, Germany.  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 3-343.  
 RX MEDLINE=97242404; PubMed=9087403;  
 RA Groll M., Ditzel U., Lowe J., Stock D., Bochtler M., Bartunik H.D.,  
 RA Huber R.;  
 RL "Structure of 20S proteasome from yeast at 2.4-A resolution.";  
 RT Nature 386:463-471(1997).  
 CC -1- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX  
 WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,  
 PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR  
 SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC  
 ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad  
 specificity.  
 CC -1- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal  
 proteolytic pathway.  
 CC -1- SUBUNIT: The yeast proteasome seems to be composed of 14 different  
 subunits which form a highly ordered ring-shaped structure.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L34348; AAA34903.1; -  
 DR EMBL: Z74780; AAA39040.1; -  
 DR PIR: S47910; S47910.  
 DR PDB: 1KVP; 15-APR-98.  
 DR MEROPS: T01.974; -  
 DR SGD: S0005398; PRE6.  
 DR InterPro: IPR000426; Proteasome\_A.  
 DR InterPro: IPR001353; Proteasome\_protease.  
 DR Pfam: PF00227; Proteasome\_1.  
 DR PROSITE: PS00388; PROTEASOME\_A; 1.  
 KW Proteasome, Hydrolyase, Protease, 3D-structure.  
 SQ SEQUENCE 254 AA; 28439 MW; 73AE63B836E618A CRC64;

Query Match 0.7%; Score 8; DB 1; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1109 SLEEVQT 1116  
 |||||

DR 198 SLEEVQT 205  
 RESULT 21  
 ID HISU\_CAMJ6 STANDARD; PRT; 256 AA.  
 AC Q46125; Q9EPH3;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Histidine-binding protein precursor (HBP) (p29).  
 GN HISU OR CDAC OR CD0734C.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M275;  
 RX MEDLINE=9735360; PubMed=9192026;  
 RA Garvis S.G., Puzon G.J., Konkel M.E.;  
 RT "Cloning, sequencing, and expression of a Campylobacter jejuni  
 periplasmic binding protein (p29) involved in histidine transport.";  
 RL Adv. Exp. Med. Biol. 412:263-264(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=72Dz/92;  
 RX MEDLINE=98055437; PubMed=9395059;  
 RA Pawelec D., Kozyniek B., Popowski J., Jaguszczyn-Krynicka E.K.;  
 RT "Cloning and characterization of a Campylobacter jejuni 72Dz/92 gene  
 encoding a 30 kDa immunopositive protein, component of the ABC  
 transport system; expression of the gene in avirulent salmonella  
 typhimurium.";  
 RL FEBS Immunol. Med. Microbiol. 19:137-150(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Paulwell T., Holtroyd S.,  
 RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrett B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 CC -1- FUNCTION: COMPONENT OF THE HIGH-AFFINITY HISTIDINE PERMEASE, A  
 BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM. THE OTHER COMPONENTS  
 ARE PROTEINS HISQ, M, AND P (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
 PROTEIN FAMILY 3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U58045; AAC35419.1; -  
 DR EMBL: Y10873; CA871823.1; -  
 DR HSSP: P31982; IHS1.  
 DR InterPro: IPR001311; SBP\_glu\_receptor.  
 DR InterPro: IPR001638; SBP\_bac\_3.  
 DR Pfam: PF00497; SBP\_bac\_3; 1.  
 DR ProDom: PD000500; Ton\_glu\_receptor; 1.  
 DR SMART: SM00062; PBP; 1.  
 DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 DR PROSITE: PS01039; SBP\_BACTERIAL\_3; 1.

KW Transport; Amino-acid transport; Signal; Membrane; Lipoprotein;  
 Complete proteome.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 256 HISTIDINE-BINDING PROTEIN.  
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT CONFLICT 4 8 FLTAIF -> ILSTA (IN REF. 3).  
 FT CONFLICT 12 13 FT -> LV (IN REF. 3).  
 FT CONFLICT 18 38 VACONTERENASNEANTITLT -> GACSDSKNKSNAVSE  
 (IN REF. 3).  
 FT CONFLICT 52 55 FMOD -> YKEN (IN REF. 3).  
 FT CONFLICT 66 66 I -> V (IN REF. 3).  
 FT CONFLICT 76 76 E -> A (IN REF. 3).  
 FT CONFLICT 93 93 S -> S (IN REF. 3).  
 FT CONFLICT 228 228 D -> N (IN REF. 3).  
 SQ SEQUENCE 256 AA; 28531 MW; 2E3B34DFCB92CB29 CRC64;  
 Query Match 0.7%; Score 8; DB 1; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 922 GTLOENTA 929  
 DB 151 GTLOENTA 158  
 RESULT 22  
 PSAT SCHPO STANDARD; PRT; 259 AA.  
 ID PSAT SCHPO  
 AC 010329;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable proteasome subunit alpha type 7 (EC 3.4.25.1).  
 GN SPBC106.16.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OC NCBI\_Taxid=4896;  
 OX NCBI\_Taxid=4896;  
 RN [1]  
 RM SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Stammers M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grynopreux B.,  
 RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Meller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wandut R., Purrelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzada J.L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpkovski G.V., Useery D., Barrett B.G., Nurse P.;  
 RA "The genome sequence of Schizosaccharomyces pombe";  
 RA Nature 415:871-880 (2002).  
 CC -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX  
 WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,  
 PHE, ITR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR  
 SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC

CC ACTIVITY (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad  
 CC specificity.  
 CC -!- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal  
 CC proteolytic pathway.  
 CC -!- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL  
 CC SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; AL110295; CAB53732.1; -  
 CC DR HSSP; P40303; LRP.  
 CC DR MEROPS; T01.974; -  
 CC DR InterPro; IPR000426; Proteasome\_A.  
 CC DR InterPro; IPR001353; Proteasome.  
 CC DR Pfam; PF00227; proteasome; 1.  
 CC DR PROSITE; PS00388; PROTEASOME\_A; 1.  
 CC KW Proteasome; Hydrolase; Protease.  
 CC SQ SEQUENCE 259 AA; 28277 MW; 6F36070FEF531380 CRC64;  
 Query Match 0.7%; Score 8; DB 1; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1109 SLEVVQRT 1116  
 DB 194 SLEVVQRT 201  
 RESULT 23  
 COMT RAT STANDARD; PRT; 264 AA.  
 ID COMT RAT  
 AC P22734;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Catechol O-methyltransferase, membrane-bound form (EC 2.1.1.6)  
 DE (MB-COMT) [Contains: Catechol O-methyltransferase, soluble form  
 DE (S-COMT)].  
 GN COMT.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_Taxid=10116;  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RM SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=94107221; PubMed=8280056;  
 RA Tenhunen J., Ullmanen I.;  
 RT "Production of rat soluble and membrane-bound catechol O-  
 RT methyltransferase forms from bifunctional mRNAs";  
 RL Biochem. J. 296:595-600 (1993).  
 CC [2]  
 CC SEQUENCE OF 11-264 FROM N.A.  
 CC MEDLINE=91033034; PubMed=2227437;  
 CC Salminen M., Lundstroem K., Tiliemann C., Savolainen R., Kalkkinen N.,  
 CC Ullmanen I.;  
 CC "Molecular cloning and characterization of rat liver catechol-O-  
 CC methyltransferase";  
 CC Nature 339:241-247 (1990).  
 CC [3]  
 CC SEQUENCE OF 1-10 FROM N.A. AND CHARACTERIZATION OF THE TWO FORMS.  
 CC MEDLINE=9211472; PubMed=1765063;  
 CC Ullmanen I., Lundstroem K.;

RT "Cell-free synthesis of rat and human catechol O-methyltransferase.  
 RT Insertion of the membrane-bound form into microsomal membranes in  
 RT vitro." ;  
 RL Eur. J. Biochem. 202:1013-1020 (1991).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF SOLUBLE FORM.  
 RX MEDLINE=94173341; PubMed=6127373;  
 RA Vldgren J., Svensson L.A., Liljas A.;  
 RT "Crystal structure of catechol O-methyltransferase." ;  
 RL Nature 368:354-358 (1994).  
 CC -1- FUNCTION: CATALYZES THE O-METHYLATION AND THEREBY THE  
 CC INACTIVATION, OF CATECHOLAMINE NEUROTRANSMITTERS AND CATECHOL  
 CC HORMONES. ALSO SHORTENS THE BIOLOGICAL HALF-LIVES OF CERTAIN  
 CC NEUROACTIVE DRUGS, LIKE L-DOPA, ALPHA-METHYL DOPA AND  
 CC ISOPROTERENOL.  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + catechol = S-  
 CC adenosyl-L-homocysteine + guaiacol.  
 CC -1- COFACTOR: MAGNESIUM.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (isoform S-COMT). Type II  
 CC membrane protein (isoform MB-COMT).  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a membrane-bound form/MB-COMT  
 CC (shown here) and a soluble form/S-COMT; are produced by  
 CC alternative initiation.  
 CC -1- PTM: The N-terminus is blocked.  
 CC -1- SIMILARITY: TO OTHER MAMMALIAN CATECHOL-O-METHYLTRANSFERASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z12651; CAI78276.1; -  
 DR EMBL; M60754; AAA40882.1; ALT\_INIT.  
 DR EMBL; M60753; AAA40881.1; ALT\_INIT.  
 DR PIR; J00787; J00787.  
 DR PIR; S22090; S22090.  
 DR PDB; 1VID; 11-JUL-96.  
 DR InterPro; IPR002935; Methyltransf\_3.  
 DR InterPro; IPR000051; SAM\_bind.  
 DR Pfam; PF01596; Methyltransf\_3; 1.  
 DR Transfaser; Methyltransferase; Neurotransmitter degradation;  
 DR Catecholamine metabolism; Transmembrane; Signal-anchor; Magnesium;  
 DR Alternative initiation; 3D-structure.  
 FT CHAIN 1 264 CATECHOL-O-METHYLTRANSFERASE, MEMBRANE-  
 FT BOUND ISOFORM.  
 FT CHAIN 45 264 CATECHOL-O-METHYLTRANSFERASE, SOLUBLE  
 FT ISOFORM.  
 FT INIT MET 44 44 FOR SOLUBLE ISOFORM.  
 FT TRANSMEM 3 19 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT ACT\_SITE 187 187  
 FT ACT\_SITE 242 242  
 FT METAL 184 184 MAGNESIUM.  
 FT METAL 212 212 MAGNESIUM.  
 FT METAL 213 213 MAGNESIUM.  
 SO SEQUENCE 264 AA; 29597 MW; F535DFP9C062854 CRC64;  
 QY Query Match 0.7%; Score 8; DB 1; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 9.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1133 GLLLLALL 1140  
 DB 9 GLLLLALL 16  
 RESULT 24  
 ID ALSK\_ECOLI STANDARD; PRT; 309 AA.  
 AC P32718;

DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE D-allose kinase (EC 2.7.1.55) (Allokinase).  
 GN ALSK OR B4084.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94089392; PubMed=8265357;  
 RA Blatner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,  
 RA Daniels D.L.;  
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the  
 RT region from 89.2 to 92.8 minutes." ;  
 RL Nucleic Acids Res. 21:5408-5417 (1993).  
 RN [2]  
 RP FUNCTION.  
 RC STRAIN=K12;  
 RX MEDLINE=98062191; PubMed=9401019;  
 RA Kim C., Song S., Park C.;  
 RT "The D-allose operon of Escherichia coli K-12." ;  
 RL J. Bacteriol. 179:7631-7637 (1997).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-allose = ADP + D-allose 6-phosphate.  
 CC -1- PATHWAY: D-ALLOSE METABOLISM. NOT ESSENTIAL FOR THIS PATHWAY.  
 CC -1- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLAR) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U00006; AAC43178.1; -  
 DR EMBL; AE000482; AAC77045.1; -  
 DR Ecogene; EGI1956; alek.  
 DR InterPro; IPR000600; ROK\_family.  
 DR Pfam; PF00480; ROK; 1.  
 DR PROSITE; PS01125; ROK; 1.  
 DR Transfaser; Kinase; Carbohydrate metabolism; Complete proteome.  
 DR SEQUENCE 309 AA; 33821 MW; 7C655FD2E5A8F2E CRC64;  
 QY Query Match 0.7%; Score 8; DB 1; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 632 GAQGAAIL 639  
 DB 288 GAQGAAIL 295  
 RESULT 25  
 ID NSMA\_HUMAN STANDARD; PRT; 423 AA.  
 AC NSMA\_HUMAN  
 AC 060906; O9BWR3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE sphingomyelin phosphodiesterase 2 (EC 3.1.4.12) (Neutral  
 DE sphingomyelinase) (NSMase) (N-SMase) (Lyso-platelet activating factor-  
 DE phospholipase C) (Lyso-PAF-PLC).  
 GN SMPD2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND TISSUE SPECIFICITY.  
 RX MEDLINE=98188255; PubMed=9520418;

RA Tomiuk S., Hofmann K., Nix M., Zumbansen M., Stoffel W.;  
 RT "Cloned mammalian neutral sphingomyelinase: functions in sphingolipid  
 RT signaling?";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:3638-3643 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP CHARACTERIZATION..  
 RX MEDLINE=20076490; PubMed=10608884;  
 RA Sawai H., Dome N., Nagan N., Hannun Y.A.;  
 RT "Function of the cloned putative neutral sphingomyelinase as  
 RT lyso-platelet activating factor-phospholipase C.";  
 RL J. Biol. Chem. 274:38331-38339 (1999).  
 CC -!- FUNCTION: Converts sphingomyelin to ceramide. Hydrolyze 1-acyl-2-  
 CC lyso-sn-glycerol-3-phosphocholine (lyso-PC) and 1-O-alkyl-2-lyso-  
 CC sn-glycerol-3-phosphocholine (lyso-platelet activating factor). The  
 CC physiological substrate seems to be lyso-PAF.  
 CC -!- CATALYTIC ACTIVITY: Sphingomyelin + H(2)O = N-acylsphingosine +  
 CC choline phosphate.  
 CC -!- COFACTOR: Magnesium.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- MISCELLANEOUS: THIS PROTEIN HAS AN OPTIMUM PH OF 6.5-7.5.  
 CC -!- SIMILARITY: BELONGS TO THE NEUTRAL SPHINGOMYELINASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: AJ222801; CA10995.1; -;  
 DR EMBL: BC000038; AA000038.1; -;  
 DR Genew; HGNC:11121; SMPD2.  
 DR MIM: 603498; -;  
 DR InterPro: IPR005135; Exo\_endo\_phos.  
 DR Pfam: PF03372; Exo\_endo\_phos.  
 DR HydroLase; Transmembrane; Magnesium.  
 KM HydroLase; Transmembrane; Magnesium.  
 FT TRANSMEM 330 350  
 FT TRANSMEM 354 374  
 FT METAL 49 49  
 FT SITE 180 180  
 FT ACT SITE 272 272  
 FT CONFLICT 3 3  
 FT SEQUENCE 423 AA; 47592 MW; 05252A923B363171 CRC64;  
 SQ  
 Query Match 0.7%; Score 8; DB 1; Length 423;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1133 GLLALL 1140  
 DB 334 GLLALL 341  
 RESULT 26  
 CPBK\_MOUSE STANDARD; PRT; 470 AA.  
 AC 062357;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 2B20 (EC 1.14.14.1) (CYPIB20) (P24) (Fragment).  
 GN CYP2B20.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 OK

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=96428606; PubMed=8831708;  
 RA Damon M., Fautrel A., Marc N., Guillozo A., Corcos L.;  
 RT "Isolation of a new mouse cDNA clone: hybrid form of cytochrome P450  
 RT 2b10 and NADPH-cytochrome P450 oxidoreductase.";  
 RL Biochem. Biophys. Res. Commun. 226:900-905 (1996).  
 CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
 CC MONOOXYGENASES. IN LIVER MICROSOMES THIS ENZYME IS INVOLVED IN AN  
 CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
 CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
 CC ACIDS, AND XENOBIOTICS.  
 CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
 CC oxidized flavoprotein + H(2)O.  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound Endoplasmic reticulum.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE LIVER AS WELL AS IN KIDNEY,  
 CC LUNG AND INTESTINE.  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: X99715; CA68051.1; ALT\_INIT.  
 DR HSSP; P00179; 1D76.  
 DR MGD; MGI:1202389; Cyp2b20.  
 DR InterPro: IPR001128; Cytochrome\_P450.  
 DR Pfam: PF00067; P450.1  
 DR PROSITE; PS00086; CYTOCHROME\_P450.1.  
 KM Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
 KM Microsome; Endoplasmic reticulum; Phosphorylation.  
 FT MOD RES 128 128  
 FT BINDING 436 436  
 FT NON TER 470 470  
 FT HEME.  
 FT SEQUENCE 470 AA; 53357 MW; 8B9CF3E2E622642 CRC64;  
 SQ  
 Query Match 0.7%; Score 8; DB 1; Length 470;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1134 LLLALL 1141  
 DB 6 LLLALL 13  
 RESULT 27  
 CPB1\_RAT STANDARD; PRT; 491 AA.  
 ID CPB1\_RAT  
 AC P00176;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 2B1 (EC 1.14.14.1) (CYP1B1) (P450-B) (P450-PB1 and  
 DE P450-PB2) (P450-LM2).  
 GN CYP2B1 OR CYP2B-1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxId=10116;  
 OK  
 RN [1]  
 RP SEQUENCE OF 6-491 FROM N.A. (ISOZYMES PB1 AND PB2).  
 RX MEDLINE=82222224; PubMed=6953431;  
 RA Fujii-Kuriyama Y., Mizukami Y., Kawajiri K., Sogawa K., Muramatsu M.;  
 RT "Primary structure of a cytochrome P-450: coding nucleotide sequence  
 RT of phenobarbital-inducible cytochrome P-450 cDNA from rat liver.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:2793-2797 (1982).  
 RN [2]  
 RP REVISIONS TO 166; 292 AND 378 (ISOZYMES PB1 AND PB2).  
 OK



RA Fujii-Kuriyama Y., Mizukami Y., Kawajiri K., Sogawa K., Muramatsu M.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:5443-5443(1982).  
 RN [3]  
 RP SEQUENCE OF 1-22.  
 RX MEDLINE=7919411; Pubmed=109438;  
 RA Botelho L.H., Ryan D.E., Levin W.;  
 RT "Amino acid compositions and partial amino acid sequences of three  
 highly purified forms of liver microsomal cytochrome P-450 from rats  
 treated with polychlorinated biphenyls, phenobarbital, or 3-  
 methylcholanthrene";  
 RT J. Biol. Chem. 254:5635-5640(1979).  
 RL [4]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=90059885; Pubmed=2583091;  
 RA Pyerin W., Taniguchi H.;  
 RT "Phosphorylation of hepatic phenobarbital-inducible cytochrome  
 P-450";  
 RL EMO J. 8:3003-3010(1989).  
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
 MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
 NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
 OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
 ACIDS, AND XENOBIOTICS.  
 CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
 oxidized flavoprotein + H(2)O.  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.  
 CC -1- INDUCTION: BY PHENOBARBITAL.  
 CC -1- PTM: PHOSPHORYLATION IS ACCOMPANIED BY A DECREASE IN ENZYME  
 ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: J00719; AAA41024.1; -;  
 DR EMBL: M37134; AAC42028.1; -;  
 DR PIR: A00176; Q4RTPB.  
 DR HSSP: P00179; I076.  
 DR InterPro: IPR001128; Cytochrome\_P450.  
 DR Pfam: PF00067; P450; 1.  
 DR PRINTS: PR00385; P450.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
 KM Microsome; Endoplasmic reticulum; Phosphorylation.  
 FT MOD RES 128 128 PHOSPHORYLATION (BY PKA).  
 FT BINDING 436 436 HEME.  
 FT VARIANT 303 303 S -> G (IN ISOZYME PB2).  
 FT VARIANT 321 322 AE -> TV (IN ISOZYME PB2).  
 FT VARIANT 337 337 L -> P (IN ISOZYME PB2).  
 FT VARIANT 339 339 T -> S (IN ISOZYME PB2).  
 FT VARIANT 344 344 S -> T (IN ISOZYME PB2).  
 SQ SEQUENCE 491 AA; 55933 MW; 74615501AD5497DD CRC64;

Query Match 0.7%; Score 8; DB 1; Length 491;  
 Best Local Similarity 100.0%; Pred.No.15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLALLLV 1141  
 |||||  
 Db 6 LLLALLLV 13

RESULT 28  
 CPB2 RAT STANDARD; PRT; 491 AA.  
 AC PC4157; 064582;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 2B2 (EC 1.14.14.1) (CYP1B2) (P450E) (P450 PB4).  
 GN CYP2B2 OR CYP2B-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83247397; Pubmed=6306654;  
 RA Mizukami Y., Sogawa K., Suwa Y., Muramatsu M., Fujii-Kuriyama Y.;  
 RT "Gene structure of a phenobarbital-inducible cytochrome P-450 in rat  
 liver";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:3958-3962(1983).  
 RN [2]  
 RP SEQUENCE.  
 RX MEDLINE=86059379; Pubmed=3877725;  
 RA Frey A.B., Waxman D.J., Kreibich G.;  
 RT "The structure of phenobarbital-inducible rat liver cytochrome P-450  
 isoenzyme PB-4. Production and characterization of site-specific  
 antibodies";  
 RL J. Biol. Chem. 260:15253-15265(1985).  
 RN [3]  
 RP SEQUENCE OF 168-491 FROM N.A.  
 RX MEDLINE=84159487; Pubmed=6689485;  
 RA Phillips I.R., Shephard E.A., Ashworth A., Rabin B.R.;  
 RT "Cloning and sequence analysis of a rat liver cDNA coding for a  
 phenobarbital-inducible microheterogenous cytochrome P-450 variant:  
 regulation of its messenger level by xenobiotics";  
 RL Gene 26:41-52(1983).  
 RN [4]  
 RP SEQUENCE OF 281-491 FROM N.A.  
 RX MEDLINE=83291091; Pubmed=6688421;  
 RA Kumar A., Raphael C., Adesnik M.;  
 RT "Cloned cytochrome P-450 cDNA. Nucleotide sequence and homology to  
 multiple phenobarbital-induced mRNA species";  
 RL J. Biol. Chem. 258:11280-11284(1983).  
 RN [5]  
 RP ERRATUM.  
 RA Kumar A., Raphael C., Adesnik M.;  
 RL J. Biol. Chem. 259:6039-6039(1984).  
 RN [6]  
 RP SEQUENCE OF 323-431 FROM N.A.  
 RX MEDLINE=86205943; Pubmed=3458196;  
 RA Atchison M.L., Adesnik M.;  
 RT "Gene conversion in a cytochrome P-450 gene family";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:2300-2304(1986).  
 RN [7]  
 RP SEQUENCE OF 385-491 FROM N.A.  
 RX MEDLINE=84153837; Pubmed=6322758;  
 RA Affolter M., Anderson A.;  
 RT "Segmental homologies in the coding and 3' non-coding sequences of  
 rat liver cytochrome P-450e and P-450b cDNAs and cytochrome  
 P-450e-like genes";  
 RL Biochem. Biophys. Res. Commun. 118:655-662(1984).  
 RN [8]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX MEDLINE=88273074; Pubmed=2839467;  
 RA Hashimoto T., Matsunoto T., Nishizawa M., Kawabata S.,  
 RA Morohashi K., Handa S., Omura T.;  
 RT "A mutant rat strain deficient in induction of a  
 phenobarbital-inducible form of cytochrome P-450 in liver  
 microsomes";  
 RL J. Biochem. 103:487-492(1988).  
 RN [9]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=90059885; Pubmed=2583091;  
 RA Pyerin W., Taniguchi H.;  
 RT "Phosphorylation of hepatic phenobarbital-inducible cytochrome  
 P-450";  
 RL EMO J. 8:3003-3010(1989).  
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
 MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN



```

CC MAMPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC ACIDS, AND XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- INDUCTION: BY PHENOBARBITAL.
CC -1- PTM: PHOSPHORYLATION IS ACCOMPANIED BY A DECREASE IN ENZYME
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, J00728; AAA41056.1; -.
CC DR EMBL, J00720; AAA41056.1; JOINED.
CC DR EMBL, J00721; AAA41056.1; JOINED.
CC DR EMBL, J00722; AAA41056.1; JOINED.
CC DR EMBL, J00723; AAA41056.1; JOINED.
CC DR EMBL, J00724; AAA41056.1; JOINED.
CC DR EMBL, J00725; AAA41056.1; JOINED.
CC DR EMBL, J00726; AAA41056.1; JOINED.
CC DR EMBL, K00996; AAA41029.1; -.
CC DR EMBL, K01626; AAA41037.1; -.
CC DR EMBL, K01721; AAA41026.1; -.
CC DR EMBL, D00250; BAA00181.1; -.
CC DR EMBL, M13234; AAA41057.1; -.
CC DR PIR, A00177; O4RP22.
CC DR PIR, A21872; A21872.
CC DR HSSP, P00179; 1DT6.
CC
CC InterPro: IPR001128; Cytochrome_P450.
CC DR Pfam, PF00067; P450; 1.
CC DR PROSITE, PS00086; CYTOCHROME_P450; 1.
CC
CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
CC Microsome; Endoplasmic reticulum; Phosphorylation.
CC MOD_RES 128 128 PHOSPHORYLATION (BY PKA).
CC FT BINDING 436 436 HEME.
CC FT CONFLICT 292 292 L -> P (IN REF. 2).
CC FT CONFLICT 321 321 T -> A (IN REF. 2 AND 4).
CC FT CONFLICT 332 332 E -> V (IN REF. 1).
CC FT CONFLICT 438 438 G -> D (IN REF. 4).
CC FT CONFLICT 444 444 N -> K (IN REF. 3).
CC FT CONFLICT 473 473 K -> M (IN REF. 1).
CC FT CONFLICT 476 476 G -> D (IN REF. 2 AND 4).
CC SQ SEQUENCE 491 AA; 55932 MW; 00CB6B937FDD44BC CRC64;
CC
CC Query Match 0.7%; Score 8; DB 1; Length 491;
CC Best Local Similarity 100.0%; Pred. No. 15;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
Qy 1134 LLLALLLV 1141
Db 6 LLLALLLV 13

```

```

CC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_Taxid=10090;
CC [1]
CC
CC SEQUENCE FROM N.A.
CC RX MEDLINE=89118235; Pubmed=3219345;
CC RA Noshiro M., Iakso M., Kawajiri K., Negishi M.;
CC "Rip locus: regulation of female-specific isozyme (I-P-450(16 alpha)
CC of testosterone 16 alpha-hydroxylase in mouse liver, chromosome
CC localization, and cloning of P-450 cDNA.";
CC Biochemistry 27:6434-6443(1988).
CC
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC MAMPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC ACIDS, AND XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M21856; AAA40425.1; -.
CC DR PIR, B31047; B31047.
CC DR HSSP, P00179; 1DT6.
CC DR MGD, MGI:88598; Cyp2b10.
CC
CC InterPro: IPR001128; Cytochrome_P450.
CC DR Pfam, PF00067; P450; 1.
CC DR PROSITE, PS00086; CYTOCHROME_P450; P450_NEG.
CC
CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
CC Microsome; Endoplasmic reticulum; Phosphorylation.
CC MOD_RES 128 128 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
CC FT BINDING 445 445 HEME.
CC SQ SEQUENCE 500 AA; 56743 MW; P66A00DD0FBA94 CRC64;
CC
CC Query Match 0.7%; Score 8; DB 1; Length 500;
CC Best Local Similarity 100.0%; Pred. No. 15;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
Qy 1134 LLLALLLV 1141
Db 6 LLLALLLV 13

```

```

CC FIBRINOGEN, PLASMINOGEN, PROTHROMBIN, THROMBOSPONDIN AND
CC VITRONECTIN. IT RECOGNIZES THE SEQUENCE R-G-D IN A WIDE ARRAY OF
CC LIGANDS. IT RECOGNIZES THE SEQUENCE H-H-L-G-G-A-K-O-A-G-D-V IN
CC FIBRINOGEN GAMMA CHAIN. FOLLOWING ACTIVATION INTEGRIN ALPHA-
CC IIB/BETA-3 BRINGS ABOUT PLATELET/PLATELET INTERACTION THROUGH
CC BINDING OF SOLUBLE FIBRINOGEN. THIS STEP LEADS TO RAPID PLATELET
CC AGGREGATION WHICH PHYSICALLY PLUGS RUPTURED ENDOTHELIAL CELL
CC SURFACE.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC DISULFIDE BOND. ALPHA-IIB ASSOCIATES WITH BETA-3.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 FG-GAP REPEAT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L12233; AAA65936.1; -.
CC HSSP; P06756; 1JY2.
CC InterPro; IPR000413; Integrin_alpha.
CC Pfam; PF00357; Integrin_A; 1.
CC DR Pfam; PF01839; FG-GAP; 1.
CC SMART; SM00191; Int_alpha; 1.
CC DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
CC KM Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
CC platelet; Calcium.
CC FT NON TER 1 1
CC FT CHAIN <1 467 INTEGRIN ALPHA-IIB HEAVY CHAIN.
CC FT CHAIN 468 604 INTEGRIN ALPHA-IIB LIGHT CHAIN.
CC FT DOMAIN <1 558 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 559 584 POTENTIAL.
CC FT DOMAIN 585 604 CYTOPLASMIC (POTENTIAL).
CC FT REPEAT 10 62 FG-GAP.
CC FT CA BIND 22 30 POTENTIAL.
CC FT SITE 587 591 GFFKR MOTIF.
CC FT DISULFID 69 80 BY SIMILARITY.
CC FT DISULFID 86 141 BY SIMILARITY.
CC FT DISULFID 198 204 BY SIMILARITY.
CC FT DISULFID 270 283 BY SIMILARITY.
CC FT DISULFID 422 476 INTERCHAIN (BY SIMILARITY).
CC FT DISULFID 481 486 BY SIMILARITY.
CC FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 604 AA; 66065 MW; 0B13BD1BD9E37F88 CRC64;

```

```

Query Match 0.7%; Score 8; DB 1; Length 604;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1130 VLGGLLLL 1137
Db 569 VLGGLLLL 576

```

```

RESULT 31
ZNF74 HUMAN STANDARD; PRT; 643 AA.
AC 016587; Q9UF05; Q9UF06; Q9UF07;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ZNF74.
DE Zinc finger protein 74.
GN ZNF74.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

OX NCBI_TaxId=9606;
RN [1] SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Brain;
RC MEDLINE=96279060; PubMed=8663113;
RA Gordin B., Bazinet M., Aubry M.;
RT "The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein
RT tightly associated with the nuclear matrix.";
RL J. Biol. Chem. 271:15458-15467(1996).
RN [2] SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=94093543; PubMed=8268910;
RA Aubry M., Desmazes S., Desmazes C., Aikem M., Aurias M., Julien J.-P.,
RA Rouleau G.A.;
RT "Isolation of a zinc finger gene consistently deleted in DisGeorge
RT syndrome.";
RL Hum. Mol. Genet. 2:1583-1587(1993).
RN [3] SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RA Aubry M., Cole F.;
RT "Alternative promoter usage and splicing of ZNF74 gene.";
RL Submitted (JUN-1998) to the EMBL/Genbank/DBSJ databases.
RN [4] VARIANTS LYS-117 AND 622-ASN-PHE-623.
RX MEDLINE=21563124; PubMed=11705709;
RA Takase K., Ohtsuki T., Migita O., Toru M., Inada T.,
RA Yamakawa-Kobayashi K., Arinami T.;
RT "Association of ZNF74 gene genotypes with age-at-onset of
RT schizophrenia.";
RL Schizophrenia. Res. 52:161-165(2001).
CC -1- FUNCTION: MAY PLAY A ROLE IN RNA METABOLISM.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2 (SHOWN HERE), 3 AND 4; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE FETAL BRAIN.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X92715; CAA63379.1; -.
CC EMBL; X71623; CAA50632.1; -.
CC DR EMBL; AF072567; AAF21777.1; -.
CC DR EMBL; AF072567; AAF21777.1; JOINED.
CC DR EMBL; AF072567; AAF21777.1; JOINED.
CC DR EMBL; AF072567; AAF21778.1; -.
CC DR EMBL; AF072567; AAF21778.1; JOINED.
CC DR EMBL; AF072567; AAF21778.1; JOINED.
CC DR EMBL; AF072567; AAF21779.1; -.
CC DR EMBL; AF072567; AAF21779.1; JOINED.
CC DR EMBL; AF072567; AAF21779.1; JOINED.
CC DR EMBL; AF072567; AAF21780.1; -.
CC DR EMBL; AF072567; AAF21780.1; JOINED.
CC DR HSSP; P07248; 1PAA.
CC DR GeneW; HGNC:1144; ZNF74.
CC MIM; 194548; -.
CC InterPro; IPR001909; KRAB.
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; Zf-C2H2; 12.
CC Pfam; PF01352; KRAB; 1.
CC PRINTS; PR00048; ZINCFINGER.
CC PRODOM; PD000003; Znf_C2H2; 9.
CC SMART; SM00349; KRAB_1.
CC SMART; SM00355; Znf_C2H2; 12.
CC DR PROSITE; PS50805; KRAB; 1.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.

```



```

FT CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 942 AA; 102387 MW; 93B300B52FF549DE CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1128 GSVLGGLL 1135
   |||||
Db 486 GSVLGGLL 493

RESULT 33
VP2_BTIVV STANDARD; PRT; 959 AA.
AC 006598;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer capsid protein VP2.
GN S2.
OS Blueongue virus (serotype 3 / isolate South Africa-vaccine).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=36424;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021485; PubMed=2171239;
RA Gould A.R., Pritchard L.I.;
RT "Relationships amongst blueongue viruses revealed by comparisons of
RL capsid and outer coat protein nucleotide sequences.";
RL Virus Res. 17:31-52(1990).
CC -1- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP5)
CC MAJOR CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE
CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.
CC -1- SIMILARITY: BELONGS TO THE REOVIRUSES VP2 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X55801; CAA39323.1; -
DR PIR: B60017; B60017.
DR InterPro: IPR001742; Orf1_VP2.
DR Pfam: PF00898; Orf1_VP2; 1.
DR ProDom: PD002938; Orf1_VP2; 1.
DR Coar protein.
SQ SEQUENCE 959 AA; 112163 MW; 41B230E31803588F CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 EKREKLE 1166
   |||||
Db 798 EKREKLE 805

RESULT 34
ITAB MOUSE STANDARD; PRT; 1033 AA.
ID ITAB_MOUSE
AC 090200; 0922M0; Q64228;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-IIB precursor (Platelet membrane glycoprotein IIB)
DE (GpIIb/IIIa) (CD41 antigen).
GN ITGA2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20042231; PubMed=10572112;
RA Thornton M.A., Poncz M.;
RT "Characterization of the murine platelet alphaIIb gene and encoded
RT cDNA.";
RL Blood 94:3947-3950(1999).
RN [2]
RP SEQUENCE OF 805-865 FROM N.A.
RX MEDLINE=92381049; PubMed=1512266;
RA Chen Y.Q., Gao X., Tamar J., Tang D., Grossi I.M., Chelladurai M.,
RA Kunicki T.J., Fligiel S.E., Taylor J.D., Honn K.V.;
RT "Identification of the alpha IIb beta 3 integrin in murine tumor
RT cells.";
RL J. Biol. Chem. 267:17314-17320(1992).
RN [3]
RP SEQUENCE OF 790-1022 FROM N.A.
RA Rout U.K., Armand D.R.;
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-484 FROM N.A.
RX MEDLINE=20179880; PubMed=10713093;
RA Puzon-McLaughlin W., Kamata T., Takada Y.;
RT "Multiple discontinuous ligand-mimetic antibody binding sites define a
RT ligand binding pocket in integrin alphaIIb beta3.";
RL J. Biol. Chem. 275:7795-7802(2000).
CC -1- FUNCTION: INTEGRIN ALPHA-IIIB/BETA-3 IS A RECEPTOR FOR FIBRONECTIN,
CC FIBRINOGEN, PLASMINOGEN, PROTHROMBIN, THROMBOSPONDIN AND
CC VITRONECTIN. IT RECOGNIZES THE SEQUENCE R-G-D IN A WIDE ARRAY OF
CC LIGANDS. IT RECOGNIZES THE SEQUENCE H-H-L-G-G-A-K-Q-A-G-D-V IN
CC FIBRINOGEN GAMMA CHAIN. FOLLOWING ACTIVATION INTEGRIN ALPHA-
CC IIB/BETA-3 BRINGS ABOUT PLATELET/PLATELET INTERACTION THROUGH
CC BINDING OF SOLUBLE FIBRINOGEN. THIS STEP LEADS TO RAPID PLATELET
CC AGGREGATION WHICH PHYSICALLY PLUGS RUPTURED ENDOTHELIAL CELL
CC SURFACE.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC DISULFIDE BOND. ALPHA-IIIB ASSOCIATES WITH BETA-3.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF169829; AAF06996.1; -
DR EMBL: AF170316; AAD58216.1; -
DR EMBL: S43388; AAB23054.2; -
DR EMBL: AF045019; AAD02339.1; -
DR EMBL: AF166384; AAF43997.1; -
DR HSSP: P06756; 1JY2.
DR MGD: MGI:96601; Itga2b.
DR InterPro: IPR00413; Integrin_alpha.
DR Pfam: PF00357; Integrin_A; 1.
DR Pfam: PF01839; FG-GAP; 5.
DR PRINTS: P01185; INTEGRINA.
DR SMART: SM00191; Int_alpha; 5.
DR PROSITE: PS00242; INTEGRIN_ALPHA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium.
FT SIGNAL 1 31 BY SIMILARITY.
FT CHAIN 32 1033 INTEGRIN ALPHA-IIIB.
FT CHAIN 32 1033 INTEGRIN ALPHA-IIIB HEAVY CHAIN (BY
FT CHAIN 32 1033 SIMILARITY).
FT CHAIN ? 1033 INTEGRIN ALPHA-IIIB LIGHT CHAIN (BY
FT SIMILARITY).

```

```

FT DOMAIN 32 988 EXTRACELLULAR (POTENTIAL).
FT TRAMSEM 989 1014 POTENTIAL.
FT DOMAIN 1015 1033 CYTOPLASMIC (POTENTIAL).
FT REPEAT 47 107 FG-GAP 1.
FT REPEAT 262 315 FG-GAP 2.
FT REPEAT 262 315 FG-GAP 3.
FT REPEAT 316 383 FG-GAP 4.
FT REPEAT 384 443 FG-GAP 5.
FT REPEAT 444 496 FG-GAP 6.
FT REPEAT 273 281 FG-GAP 7.
FT CA_BIND 327 335 POTENTIAL.
FT CA_BIND 395 403 POTENTIAL.
FT CA_BIND 456 464 POTENTIAL.
FT SITE 1017 1021 GEFKR MOTIF.
FT DISULFID 87 96 BY SIMILARITY.
FT DISULFID 138 161 BY SIMILARITY.
FT DISULFID 177 197 BY SIMILARITY.
FT DISULFID 503 514 BY SIMILARITY.
FT DISULFID 520 575 BY SIMILARITY.
FT DISULFID 632 638 BY SIMILARITY.
FT DISULFID 704 717 BY SIMILARITY.
FT DISULFID 856 905 INTERCHAIN (BY SIMILARITY).
FT DISULFID 911 916 BY SIMILARITY.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 957 957 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 236 236 T->S (IN REF. 4).
FT CONFLICT 270 270 S->A (IN REF. 4).
FT CONFLICT 464 464 D->G (IN REF. 4).
FT CONFLICT 471 471 W->G (IN REF. 4).
FT CONFLICT 483 483 G->V (IN REF. 4).
FT CONFLICT 805 805 S->R (IN REF. 2 AND 3).
FT CONFLICT 834 834 P->H (IN REF. 3).
FT CONFLICT 849 849 VQ->LR (IN REF. 3).
FT CONFLICT 865 865 D->E (IN REF. 2).
FT CONFLICT 933 933 A->V (IN REF. 3).
SQ SEQUENCE 1033 AA; 112697 MW; 7F20A16C8EF2BD7 CRC64;

```

```

Query Match 0.7%; Score 8; DB 1; Length 1033;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1130 VVGGLLL 1137
DB 999 VVGGLLL 1006

```

```

RESULT 35
ITAB_HUMAN STANDARD; PRT; 1039 AA.
AC P08514; Q1443; G95366;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-IIb precursor (Platelet membrane glycoprotein IIb)
DE (GPIIb) (CD41 antigen).
OS ITGA2B OR ITGAB OR GP2B.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=8750457; PubMed=2439501;
RA Poncz M., Eisman R., Heideneich R., Silver S.M., Vilaire G.,
RA Surety S., Schwartz E., Bennett J.S.;
RT "Structure of the platelet membrane glycoprotein IIb. Homology to the
RT alpha subunits of the vitronectin and fibronectin membrane
RT receptors."
RL J. Biol. Chem. 262:8476-8482(1987).

```

```

RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=90265363; PubMed=2345548;
RA Frachet P., Uzan G., Thevenon D., Denarier E., Prandini M.H.,
RA Marguerite G.;
RT "GPIIb and GPIIa amino acid sequences deduced from human
RT megakaryocyte cDNAs."
RL Mol. Biol. Rep. 14:27-33(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=90212612; PubMed=2322558;
RA Heideneich R., Eisman R., Surety S., Delgrosso K., Bennett J.S.,
RA Schwartz E., Poncz M.;
RT "Organization of the gene for platelet glycoprotein IIb."
RL Biochemistry 29:1232-1244(1990).
RN [4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Erythrocyte;
RX MEDLINE=90277633; PubMed=2351656;
RA Bray P.F., Leung C.S.-I., Shuman M.A.;
RT "Human platelets and megakaryocytes contain alternately spliced
RT glycoprotein IIb mRNAs."
RL U. Biol. Chem. 265:9587-9590(1990).
RN [5]
RP SEQUENCE OF 32-56 AND 903-917.
RX MEDLINE=87041455; PubMed=3534886;
RA Charo I.F., Fitzgerald L.A., Steiner B., Rall S.C., Bekeart L.S.,
RA Phillips D.R.;
RT "Platelet glycoproteins IIb and IIla: evidence for a family of
RT immunologically and structurally related glycoproteins in mammalian
RT cells."
RL Proc. Natl. Acad. Sci. U.S.A. 83:8351-8355(1986).
RN [6]
RP SEQUENCE OF 392-1039 FROM N.A. (ISOFORM 1).
RX MEDLINE=8811709; PubMed=3422188;
RA Uzan G., Frachet P., Lajmanovich A., Prandini M.H., Denarier E.,
RA Duperray A., Loftus J., Ginsberg M., Plow E., Marguerite G.;
RT "cDNA clones for human platelet GPIIb corresponding to mRNA from
RT megakaryocytes and HBL cells. Evidence for an extensive homology to
RT other Arg-Gly-Asp adhesion receptors."
RL Eur. J. Biochem. 171:87-93(1988).
RN [7]
RP SEQUENCE OF 868-1039 FROM N.A. (ISOFORM 1).
RX MEDLINE=88059639; PubMed=3479442;
RA Bray P.F., Rosa J.P., Johnston G.I., Shiu D.T., Cook R.G., Lau C.,
RA Kan Y.W., McEwen R.P., Shuman M.A.;
RT "Platelet glycoprotein IIb. Chromosomal localization and tissue
RT expression."
RL J. Clin. Invest. 80:1812-1817(1987).
RN [8]
RP SEQUENCE OF 1-62 AND 1021-1039 FROM N.A.
RX MEDLINE=89025907; PubMed=2845986;
RA Prandini M.H., Denarier E., Frachet P., Uzan G., Marguerite G.;
RT "Isolation of the human platelet glycoprotein IIb gene and
RT characterization of the 5' flanking region."
RL Biochem. Biophys. Res. Commun. 156:595-601(1988).
RN [9]
RP SEQUENCE OF 487-501 AND 1026-1038.
RX MEDLINE=87101510; PubMed=3801670;
RA Hiraiwa A., Matsukage A., Shiku H., Takahashi T., Naito K., Yamada K.;
RT "Purification and partial amino acid sequence of human platelet
RT membrane glycoproteins IIb and IIla."
RL Blood 69:560-564(1987).
RN [10]
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=89374157; PubMed=2775232;
RA Calvete J.J., Henschen A., Gonzalez-Rodriguez J.;
RT "Complete localization of the intrachain disulphide bonds and the N-
RT glycoprotein IIb."
RL Biochem. J. 261:561-568(1989).
RN [11]
RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITE SER-874.

```

RX MEDLINE=93345693; PubMed=7686323;  
 RA Calvee J.U., Munitz-Diaz E.;  
 RT "Localization of an O-glycosylation site in the alpha-subunit of the  
 human platelet integrin GPIIb/IIIa involved in Baka (HPA-3a)  
 RT allonitigen expression.";  
 RL FEBS Lett. 328:30-34(1993).  
 RN [12]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 3).  
 RA MEDLINE=99025846; PubMed=9809974;  
 RT Trikha M., Cai Y., Grignon D., Honn K.V.;  
 RT "Identification of a novel truncated alphaIIb integrin.";  
 RL Cancer Res. 58:4771-4775(1998).  
 RN [13]  
 RP MUTAGENESIS OF PRO-1029 AND PRO-1030.  
 RA MEDLINE=99230329; PubMed=10212286;  
 RT Leisner T.M., Mencil-Drake J.D., Wang W., Lam S.C.;  
 RT "Bidirectional transmembrane modulation of integrin alphaIIb beta3  
 RT conformational.";  
 RL J. Biol. Chem. 274:12945-12949(1999).  
 RN [14]  
 RP VARIANT HPA-3 (BAK).  
 RA MEDLINE=90275262; PubMed=2350579;  
 RT Lyman S., Aeter R.H., Visentin G.P., Newman P.J.;  
 RT "Polymorphism of human platelet membrane glycoprotein IIb associated  
 RL Blood 75:2343-2348(1990).  
 RN [15]  
 RP VARIANT GTA ASP-273.  
 RA MEDLINE=94110321; PubMed=8282784;  
 RA Poncz M., Rifat S., Collier B.S., Newman P.J., Shattil S.J.,  
 RA Parrella T., Fortina P., Bennett J.S.;  
 RT "Glanzmann thrombasthenia secondary to a Gly273-->ASP mutation  
 RT adjacent to the first calcium-binding domain of platelet glycoprotein  
 RT IIb.";  
 RL J. Clin. Invest. 93:172-179(1994).  
 RN [16]  
 RP VARIANT GTA ASP-449.  
 RA MEDLINE=94140878; PubMed=7508443;  
 RA Wilcox D.A., Mautler J.-L., Pidar D., Newman P.J.;  
 RT "A single amino acid substitution flanking the fourth calcium binding  
 RT domain of alpha IIb prevents maturation of the alpha IIb beta 3  
 RT integrin complex.";  
 RL J. Biol. Chem. 269:4450-4457(1994).  
 RN [17]  
 RP VARIANT GTA HIS-358.  
 RA MEDLINE=95221604; PubMed=7706461;  
 RA Wilcox D.A., Paddock C.M., Lyman S., Gill J.C., Newman P.J.;  
 RT "Glanzmann thrombasthenia resulting from a single amino acid  
 RT substitution between the second and third calcium-binding domains of  
 RT GPIIb. Role of the GPIIb amino terminus in integrin subunit  
 RT association.";  
 RL J. Clin. Invest. 95:1553-1560(1995).  
 RN [18]  
 RP REVIEW ON GTA VARIANTS.  
 RA MEDLINE=95184171; PubMed=7878622;  
 RA Bray P.F.;  
 RT "Inherited diseases of platelet glycoproteins: considerations for  
 RT rapid molecular characterization.";  
 RL Thromb. Haemost. 72:492-502(1994).  
 RN [19]  
 RP VARIANT GTA PRO-778.  
 RA MEDLINE=98438330; PubMed=9763559;  
 RA Tadokoro S., Tomiyama Y., Honda S., Yamamoto N., Shiraga M.,  
 RA Kosugi S., Kanakura Y., Kurata Y., Matsuzawa Y.;  
 RT "A Glu747-->Pro substitution in the IIb subunit is responsible for a  
 RT moderate IIb beta3 deficiency in Glanzmann thrombasthenia.";  
 RL Blood 92:2750-2758(1998).  
 RN [20]  
 RP VARIANTS GTA SER-320; LYS-355 AND PRO-778.  
 RA MEDLINE=98387769; PubMed=9722314;  
 RA Ambo H., Kanata T., Honda M., Kawai Y., Oda A., Murata M., Takada Y.,  
 RA Ikeda Y.;  
 RT "Novel point mutations in the alphaIIb subunit (Phe289-->Ser,

RT Glu324-->Lys and Glu747-->Pro) causing thrombasthenic phenotypes in  
 RT four Japanese patients.";  
 RL Br. J. Haematol. 102:829-840(1998).  
 RN [21]  
 RP VARIANT GTA ALA-176 AND LEU-176.  
 RA MEDLINE=20076256; PubMed=10607701;  
 RA Basant R.B., French D.L., Vialre G., Brown D.L., Chen F.,  
 RA Collier B.S., Derick J.M., Garner T.K., Bennett J.S., Poncz M.;  
 RT "A naturally occurring mutation near the amino terminus of alphaIIb  
 RT defines a new region involved in ligand binding to alphaIIb beta3.";  
 RL Blood 95:180-188(2000).  
 RN [22]  
 RP FUNCTION: INTEGRIN ALPHA-IIb/BETA-3 IS A RECEPTOR FOR FIBRONECTIN,  
 CC FIBRINOGEN, PLASMINOGEN, PROTHROMBIN, THROMBOSPONDIN AND  
 CC VITRONECTIN. IT RECOGNIZES THE SEQUENCE R-G-D IN A WIDE ARRAY OF  
 CC LIGANDS. IT RECOGNIZES THE SEQUENCE H-H-L-G-G-A-K-Q-A-G-D-V IN  
 CC FIBRINOGEN GAMMA CHAIN FOLLOWING ACTIVATION INTEGRIN ALPHA-  
 CC IIb/BETA-3 BRINGS ABOUT PLATELET/PLATELET INTERACTION THROUGH  
 CC BINDING OF SOLUBLE FIBRINOGEN. THIS STEP LEADS TO RAPID PLATELET  
 CC AGGREGATION WHICH PHYSICALLY PLUGS RUPTURED ENDOTHELIAL CELL  
 CC SURFACE.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA  
 CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A  
 CC DISULFIDE BOND. ALPHA-IIb ASSOCIATES WITH BETA-3.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: ISOFORM 1 AND ISOFORM 2 WERE IDENTIFIED IN  
 CC PLATELETS AND MEGACARYOCYTES, BUT NOT IN RETICULOCYTES OR IN  
 CC JURKAT AND U937 WHITE BLOOD CELL LINE. ISOFORM 3 IS EXPRESSED BY  
 CC LEUCEMIA, PROSTATE ADENOCARCINOMA AND MELANOMA CELLS BUT NOT BY  
 CC PLATELETS OR NORMAL PROSTATE OR BREAST EPITHELIAL CELLS.  
 CC -1- POLYMORPHISM: POSITION 874 IS ASSOCIATED WITH PLATELET-SPECIFIC  
 CC ALLOANTIGEN HPA-3/BAK/LEK. HPA-3/BAK(A)/LEK(A) HAS ILE-874 AND

Query Match 0.7%; Score 8; DB 1; Length 1039;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1130 VLGGLLL 1137  
 Db 1004 VLGGLLL 1011

RESULT 36  
 ITAS XENLA STANDARD; PRT: 1050 AA.  
 AC 006274;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-5 precursor (Fibronectin receptor alpha subunit)  
 DE (Integrin alpha-P) (VLA-5).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95344994; PubMed=7619730;  
 RA Joce T.O., Whitaker C.A., Meng F., Desimone D.W., Grau V.,  
 RA Hausen P.;  
 RT "Integrin alpha 5 during early development of Xenopus laevis.";  
 RL Mech. Dev. 50:187-199(1995).  
 RN [2]  
 RP SEQUENCE OF 318-393 FROM N.A.  
 RA MEDLINE=94008528; PubMed=8404528;  
 RA Whitaker C.A., Desimone D.W.;  
 RT "Integrin alpha subunit mRNAs are differentially expressed in early  
 RT Xenopus embryos.";  
 RL Development 117:1239-1249(1993).  
 CC -1- FUNCTION: INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRONECTIN.  
 CC IT RECOGNIZES THE SEQUENCE R-G-D IN ITS LIGANDS.









RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.  
 RA MEDLINE=95171458; PubMed=7867070;  
 RX Lee J.O., Rieu P., Arnaout M.A., Liddington R.;  
 RT "Crystal structure of the A domain from the alpha subunit of integrin  
 RT CR3 (CD11b/CD18)";  
 RL Cell 80:631-638 (1995).  
 RM (10)  
 RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
 RP MEDLINE=96363671; PubMed=8747460;  
 RX Lee J.O., Bankston L.A., Arnaout M.A., Liddington R.C.;  
 RT "Two conformations of the integrin A-domain (I-domain): a pathway for  
 RT activation?";  
 RL Structure 3:1333-1340 (1995).  
 RM (11)  
 RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
 RP MEDLINE=968362595; PubMed=9687375;  
 RX Baldwin E.T., Sarver R.W., Bryant G.L. Jr., Curry K.A.,  
 RA Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrikson R.L.,  
 RA Horton N.C., Kelley L.L., Milder A.M., Moon J.B., Mott J.E.,  
 RA Mutchler V.T., Tomich C.S., Watenpaugh K.D., Wiley V.H.;  
 RT "Cation binding to the integrin CD11b I domain and activation model  
 RT assessment";  
 RL Structure 6:923-935 (1998).  
 RM (12)  
 RN 3D-STRUCTURE MODELING OF 17-616.  
 RP MEDLINE=98226734; PubMed=9560195;  
 RX Oxyg C., Springer T.A.;  
 RT "Experimental support for a beta-propeller domain in integrin alpha-  
 RT subunits and a calcium binding site on its lower surface";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875 (1998).  
 CC -1- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS  
 CC ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES  
 CC AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.  
 CC IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF  
 CC THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D  
 CC PEPTIDE IN CSB. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR  
 CC FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES  
 CC OF FIBRINOGEN GAMMA CHAIN.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-M  
 CC ASSOCIATES WITH BETA-2.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
 CC GRANULOCYTES.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VMFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VMFA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD11b entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC  
 DR EMBL, J03925; AAA59544.1; -  
 DR EMBL, M18044; AAA59491.1; -  
 DR EMBL, J04145; AAA59903.1; -  
 DR EMBL, S52237; AAB24821.1; -  
 DR EMBL, S52152; AAB24821.1; JOINED.  
 DR EMBL, S52153; AAB24821.1; JOINED.  
 DR EMBL, S52154; AAB24821.1; JOINED.  
 DR EMBL, S52155; AAB24821.1; JOINED.  
 DR EMBL, S52157; AAB24821.1; JOINED.  
 DR EMBL, S52159; AAB24821.1; JOINED.  
 DR EMBL, S52161; AAB24821.1; JOINED.  
 DR EMBL, S52164; AAB24821.1; JOINED.  
 DR EMBL, S52165; AAB24821.1; JOINED.  
 DR EMBL, S52167; AAB24821.1; JOINED.

DR EMBL, S52169; AAB24821.1; JOINED.  
 DR EMBL, S52170; AAB24821.1; JOINED.  
 DR EMBL, S52173; AAB24821.1; JOINED.  
 DR EMBL, S52174; AAB24821.1; JOINED.  
 DR EMBL, S52180; AAB24821.1; JOINED.  
 DR EMBL, S52181; AAB24821.1; JOINED.  
 DR EMBL, S52184; AAB24821.1; JOINED.  
 DR EMBL, S52189; AAB24821.1; JOINED.  
 DR EMBL, S52191; AAB24821.1; JOINED.  
 DR EMBL, S52192; AAB24821.1; JOINED.  
 DR EMBL, S52203; AAB24821.1; JOINED.  
 DR EMBL, S52212; AAB24821.1; JOINED.  
 DR EMBL, S52213; AAB24821.1; JOINED.  
 DR EMBL, S52216; AAB24821.1; JOINED.  
 DR EMBL, S52219; AAB24821.1; JOINED.  
 DR EMBL, S52220; AAB24821.1; JOINED.  
 DR EMBL, S52221; AAB24821.1; JOINED.  
 DR EMBL, S52222; AAB24821.1; JOINED.  
 DR EMBL, S52226; AAB24821.1; JOINED.  
 DR EMBL, M76724; AAA58410.1; -  
 DR EMBL, M84477; AAA51960.1; -  
 DR PIR, A31108; RWHULB.  
 DR PIR, A26091; A26091.  
 DR PDB, 1A8X; 17-JUN-98.  
 DR PDB, 1BHQ; 18-NOV-98.  
 DR PDB, 1BHQ; 18-NOV-98.  
 DR PDB, 1IDN; 25-NOV-98.  
 DR PDB, 1IDO; 01-AUG-96.  
 DR PDB, 1IDM; 11-JAN-97.  
 DR Genew, HGNC:6149; ITGAM.  
 DR MIM, 120980; -  
 DR InterPro, IPR00413; Integrin\_alpha.  
 DR InterPro, IPR02035; VMF\_A.  
 DR Pfam, PF00092; vwa; 1.  
 DR Pfam, PF00357; integrin\_A; 1.  
 DR Pfam, PF01839; FG-GAP; 5.  
 DR PRINTS, PR01185; INTEGRINA.  
 DR PRINTS, PR00453; VMFADOMAIN.  
 DR SMART, SM00191; Int\_alpha; 4.  
 DR SMART, SM00327; VMA; 1.  
 DR PROSITE, PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE, PS50234; VMFA; 1.  
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal; 3D-structure; Repeat; Magnesium; Calcium.  
 FT SIGNAL 1 16  
 FT CHAIN 17 1152 INTEGRIN ALPHA-M.  
 FT DOMAIN 17 1104 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1105 1128 POTENTIAL.  
 FT DOMAIN 1129 1152 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 31 84 FG-GAP 1.  
 FT REPEAT ? ? FG-GAP 2.  
 FT DOMAIN 164 350 VMFA.  
 QY 1132 GGLLLAL 1139  
 DB 1115 GGLLLAL 1122  
 RESULT 39  
 ID ITAX HUMAN STANDARD; PRT; 1163 AA.  
 AC P20702;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-X precursor (leukocyte adhesion glycoprotein p150.95  
 DE alpha chain) (leukocyte adhesion receptor p150.95) (CD11c) (Leu M5).  
 GN ITGAX OR CD11c.  
 OS Homo sapiens (Human).  
 Query Match 0.7%; Score 8; DB 1; Length 1152;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RX MEDLINE=88166645; PubMed=3327687;  
 RA Corbi A.L., Miller L.J., O'Connor K., Larson R.S., Springer T.A.;  
 RT "CDNA cloning and complete primary structure of the alpha subunit of  
 a leukocyte adhesion glycoprotein, p150,95.";  
 RL EMO J. 6:4023-4028(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=90153906; PubMed=2303426;  
 RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;  
 RT "Genomic structure of an integrin alpha subunit, the leukocyte  
 p150,95 molecule.";  
 RL J. Biol. Chem. 265:2782-2788(1990).  
 RN [3]  
 RP ERRATUM.  
 RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;  
 RL J. Biol. Chem. 265:12750-12751(1990).  
 RN [4]  
 RP SEQUENCE OF 20-43.  
 RX MEDLINE=87167596; PubMed=3549901;  
 RA Miller L.J., Wiebe M., Springer T.A.;  
 RT "Purification and alpha subunit N-terminal sequences of human Mac-1  
 and p150,95 leukocyte adhesion proteins.";  
 RL J. Immunol. 138:2381-2383(1987).  
 CC -1- FUNCTION: INTEGRIN ALPHA-X/BETA-2 IS A RECEPTOR FOR FIBRINOGEN. IT  
 CC RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN. IT MEDIATES CELL-CELL  
 CC INTERACTION DURING INFLAMMATORY RESPONSES. IT IS ESPECIALLY  
 CC IMPORTANT IN MONOCYTE ADHESION AND CHEMOTAXIS.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-X  
 CC ASSOCIATES WITH BETA-2.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
 CC GRANULOCYTES.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD11c entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11c.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M81695; AAA59180.1; -;  
 DR EMBL; Y00093; CAA68283.1; -;  
 DR EMBL; M29165; -; NOT ANNOTATED CDS.  
 DR EMBL; M29487; AAA51620.1; ALT SEQ.  
 DR EMBL; M29487; AAA51620.1; JOINED.  
 DR EMBL; M29483; AAA51620.1; JOINED.  
 DR EMBL; M29484; AAA51620.1; JOINED.  
 DR EMBL; M29485; AAA51620.1; JOINED.  
 DR EMBL; M29486; AAA51620.1; JOINED.  
 DR PIR; A35584; RWHUIC.  
 DR HSSP; P11215; 1A8X.  
 DR Genew; HGNC; 6152; ITGAX.  
 DR MIM; 151510; -;  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00037; Vwf; 1.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.

DR SMART; SM00191; Int\_alpha; 4.  
 DR SMART; SM00327; Vwf; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS00334; VFMA; 1.  
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal; Magnesium; Calcium; Repeat.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1163 INTEGRIN ALPHA-X.  
 FT DOMAIN 20 1107 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1108 1128 POTENTIAL.  
 FT DOMAIN 1129 1163 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 34 87 FG-GAP 1.  
 FT REPEAT 34 87 FG-GAP 2.  
 FT REPEAT 165 351 VFMA.  
 FT REPEAT 402 453 FG-GAP 3.  
 FT REPEAT 455 517 FG-GAP 4.  
 FT REPEAT 518 576 FG-GAP 5.  
 FT REPEAT 581 633 FG-GAP 6.  
 FT CA\_BIND 466 474 FG-GAP 7.  
 FT CA\_BIND 530 538 POTENTIAL.  
 FT CA\_BIND 593 601 POTENTIAL.  
 FT SITE 1131 1135 GPPRR MOTIF.  
 FT DISULFID 69 76 BY SIMILARITY.  
 FT DISULFID 108 126 BY SIMILARITY.  
 FT DISULFID 655 712 BY SIMILARITY.  
 FT DISULFID 771 777 BY SIMILARITY.  
 FT DISULFID 848 863 BY SIMILARITY.  
 FT DISULFID 998 1022 BY SIMILARITY.  
 FT DISULFID 1027 1032 BY SIMILARITY.  
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 332 392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 697 697 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 939 939 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 490 490 G -> A (IN REF. 2).  
 FT CONFLICT 756 756 L -> D (IN REF. 2).  
 SQ SEQUENCE 1163 AA; 12785 MW; 6C4E19CC3F82A473 CRC64;  
 Query Match 0.7%; Score 8; DB 1; Length 1163;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1132 GGLLLAL 1139  
 DB 1115 GGLLLAL 1122  
 RESULT 40  
 CA36 HUMAN STANDARD; PRT; 3176 AA.  
 ID CA36\_HUMAN  
 AC P12111; Q16501;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 3(VI) chain precursor.  
 GN COL6A3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RP TISSUE=fibroblast;  
 RX MEDLINE=90151612; PubMed=1689238;  
 RA Chu M.-L., Zhang R.-Z., Pan T.-C., Stokes D., Conway D., Kuo H.-J.,  
 RA Glaville R., Mayer U., Mann K., Deutzmann R., Timpl R.;  
 RT "Mosaic structure of globular domains in the human type VI collagen  
 alpha 3 chain: similarity to von Willebrand factor, fibronectin,  
 RT actin, salivary proteins and aprotinin type protease inhibitors.";

EMBO J. 9:385-393(1990).  
 (2)  
 RA REVISIONS.  
 RA Chu M.-L.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 2038-2373 FROM N.A.  
 RX MEDLINE=89066644; PubMed=3198591;  
 RA Chu M.-L., Conway D., Pan T.-C., Baldwin C., Mann K., Deutzmann R.,  
 RA Timpl R.;  
 RT "Amino acid sequence of the triple-helical domain of human collagen  
 RT type VI".  
 RL J. Biol. Chem. 263:18601-18606(1988).  
 RN [4]  
 RP SEQUENCE OF 2092-2157 FROM N.A.  
 RX MEDLINE=88029444; PubMed=3665927;  
 RA Chu M.-L., Mann K., Deutzmann R., Pribula-Conway D.,  
 RA Hsu-Chen C.-C., Bernard M.P., Timpl R.;  
 RT "Characterization of three constituent chains of collagen type VI by  
 RT peptide sequences and cDNA clones".  
 RL Eur. J. Biochem. 168:309-317(1987).  
 RN [5]  
 RP SEQUENCE OF 2092-2151 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=88161046; PubMed=3348212;  
 RA Weil D., Mattei M.-G., Passagge E., van Cong N., Pribula-Conway D.,  
 RA Mann K., Deutzmann R., Timpl R., Chu M.-L.;  
 RT "Cloning and chromosomal localization of human genes encoding the  
 RT three chains of type VI collagen".  
 RL Am. J. Hum. Genet. 42:435-445(1988).  
 RN [6]  
 RP SEQUENCE OF 32-236 FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=93054780; PubMed=1339440;  
 RA Zanussi S., Doliana R., Segat D., Bonaldo P., Colombatti A.;  
 RT "The human type VI collagen gene. mRNA and protein variants of the  
 RT alpha 3 chain generated by alternative splicing of an additional 5'-end  
 RT exon".  
 RL J. Biol. Chem. 267:24082-24089(1992).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 3107-3164.  
 RX MEDLINE=95182468; PubMed=7533217;  
 RA Arnoux B., Merigieu K., Saludjian P., Norris F., Norris K., Bjoern S.,  
 RA Olsen O., Petersen L., Ducruix A.;  
 RT "The 1.6 A structure of Kunitz-type domain from the alpha 3 chain of  
 RT human type VI collagen".  
 RL J. Mol. Biol. 246:609-617(1995).  
 RN [8]  
 RP STRUCTURE BY NMR OF 3102-3164.  
 RX MEDLINE=96398604; PubMed=8805527;  
 RA Zweckstetter M., Czisch M., Mayer U., Chu M.-L., Zinth W., Timpl R.,  
 RA Holak T.A.;  
 RT "Structure and multiple conformations of the kunitz-type domain from  
 RT human type VI collagen alpha3 (VI) chain in solution".  
 RL Structure 4:195-209(1996).  
 RN [9]  
 RP STRUCTURE BY NMR OF 3107-3164.  
 RX MEDLINE=97410331; PubMed=9265624;  
 RA Soerensen M.D., Bjoern S., Norris K., Olsen O., Petersen L.,  
 RA James T.L., Led J.J.;  
 RT "Solution structure and backbone dynamics of the human alpha3-chain  
 RT type VI collagen C-terminal Kunitz domain".  
 RL Biochemistry 36:10439-10450(1997).  
 RN [10]  
 RP VARIANT BM GLU-1679, AND VARIANT HIS-2831.  
 RX MEDLINE=98204804; PubMed=9536084;  
 RA Pan T.-C., Zhang R.-Z., Pericak-Vance M.A., Tandan R., Fries T.,  
 RA Stajich J.M., Viles K., Vance J.M., Chu M.-L., Speer M.C.;  
 RT "Missense mutation in a von Willebrand factor type A domain of the  
 RT alpha 3 (VI) collagen gene (COL6A3) in a family with Bethlem  
 RT myopathy".  
 RL Hum. Mol. Genet. 7:807-812(1998).  
 CC -1- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1 (VI),

ALPHA 2 (VI), AND ALPHA 3 (VI).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- DISEASE: DEFECTS IN COL6A3 ARE A CAUSE OF BETHLEM MYOPATHY (BM).  
 CC BM IS A RARE AUTOSOMAL DOMINANT PROXIMAL MYOPATHY CHARACTERIZED BY  
 CC EARLY CHILDHOOD ONSET (COMPLETE PENETRANCE BY THE AGE OF 5) AND  
 CC JOINT CONTRACTURES MOST FREQUENTLY AFFECTING THE ELBOWS AND  
 CC ANKLES.  
 CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 12 WFPA DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X52022; CAA36267.1; -;  
 DR EMBL; X06196; CAA29557.1; -;  
 DR EMBL; M20778; -; NOT ANNOTATED\_CDS.  
 DR EMBL; M27449; AAA52057.1; -;  
 DR EMBL; S49432; AAB24261.1; -;  
 DR PIR; C31953; C31952.  
 DR PDB; 1KNT; 01-NOV-94.  
 DR PDB; 2KNT; 15-MAY-97.  
 DR PDB; 1KUN; 12-NOV-97.  
 DR Genew; HGNC:2213; COL6A3.  
 DR MIM; 120250; -;  
 DR MIM; 158810; -;  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR InterPro; IPR002035; WFP A.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR Pfam; PF00092; wfa; 11.  
 DR Pfam; PF01391; Collagen; 5.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR PRODOM; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00060; FN3; 1.  
 DR SMART; SM00131; KU; 1.  
 DR SMART; SM00327; WFA; 12.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS00279; BPTI\_KUNITZ\_2; 1.  
 DR PROSITE; PS50234; WFP A; 12.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Cell adhesion; Serine protease inhibitor; Collagen;  
 KW Signal; 3d-structure; Disease mutation; Polymorphism;  
 KW Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 3176  
 FT DOMAIN 26 2038  
 FT DOMAIN 2039 2375  
 FT DOMAIN 2376 3176  
 FT DOMAIN 39 213  
 FT DOMAIN 242 419  
 FT DOMAIN 445 620  
 FT DOMAIN 639 816  
 FT DOMAIN 837 1009  
 FT DOMAIN 1029 1205  
 FT DOMAIN 1233 1404  
 FT DOMAIN 1436 1609  
 FT DOMAIN 1639 1812  
 FT DOMAIN 1838 2024  
 FT DOMAIN 2402 2581  
 FT DOMAIN 2619 2815  
 FT DOMAIN 2987 3076  
 FT DOMAIN 3107 3176  
 POTENTIAL.  
 COLLAGEN ALPHA 3 (VI) CHAIN.  
 NONHELICAL REGION.  
 TRIPLE-HELICAL REGION.  
 NONHELICAL REGION.  
 WFP A 1.  
 WFP A 2.  
 WFP A 3.  
 WFP A 4.  
 WFP A 5.  
 WFP A 6.  
 WFP A 7.  
 WFP A 8.  
 WFP A 9.  
 WFP A 10.  
 WFP A 11.  
 WFP A 12.  
 FIBRONECTIN TYPE-III.  
 BPTI/KUNITZ INHIBITOR.

```

FT SITE 2040 2042 CELL ATTACHMENT SITE.
FT SITE 2136 2138 CELL ATTACHMENT SITE.
FT SITE 2148 2150 CELL ATTACHMENT SITE.
FT SITE 2154 2156 CELL ATTACHMENT SITE.
FT SITE 2370 2372 CELL ATTACHMENT SITE.
FT ACT_SITE 3121 3122 REACTIVE BOND.
FT DISULFID 3111 3161
FT DISULFID 3120 3144
FT DISULFID 3136 3157
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2079 2079 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2331 2331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2558 2558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2677 2677 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2861 2861 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3036 3036 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 32 236 MISSING (IN ISOFORM 2).
FT VARIANT 1679 1679 G -> E (IN BM).
FT VARIANT 2831 2831 /FTID=VAR_001910.
FT VARIANT 2831 2831 D -> H.
FT CONFLICT 127 128 /FTID=VAR_001911.
FT CONFLICT 137 137 QS -> AK (IN REF. 6).
FT CONFLICT 2157 2157 R -> L (IN REF. 6).
FT CONFLICT 2157 2157 P -> R (IN REF. 4).
SQ SEQUENCE 3176 AA; 343548 MW; SDF825632298B2D CRC64;

```

Query Match Best Local Similarity 0.7%; Score 8; DB 1; Length 3176;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 IQVGLVQY 209

DB 1676 IQVGLVQY 1683

```

RESULT 41
DMS4_AGAN STANDARD; PRT; 72 AA.
AC 093224;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dermaseptin AA-3-3 precursor.
OS Agalychnis amae (Yellow-eye leaf frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Phyllomedusinae; Agalychnis.
OX NCBI_TaxID=75990;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Skin;
RX MEDLINE=98449786; PubMed=9774745;
RA Wechsclberger C.;
RT "Cloning of cDNAs encoding new peptides of the dermaseptin-family.";
RL Biochim. Biophys. Acta 1388:279-283(1998).
CC -1- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST GRAM-
CC POSITIVE AND GRAM-NEGATIVE BACTERIA. PROBABLY ACTS BY DISTURBING
CC MEMBRANE FUNCTIONS WITH ITS AMPHIPATHIC STRUCTURE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SKIN-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO THE DERMASEPTIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE DERMASEPTIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ005186; CAA06423.1; -
CC InterPro; IPR004275; Brevenin.

```

```

DR Pfam; PF03032; Brevenin; 1.
KW Antibiotic; Multigene family; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 41 POTENTIAL.
FT CHAIN 44 69 DERMASEPTIN AA-3-3.
FT PROPEP 71 72 POTENTIAL.
FT MOD_RES 69 69 AMIDATION (G-70 PROVIDE AMIDE GROUP)
SQ SEQUENCE 72 AA; 7996 MW; B16711501A43A42F CRC64;

```

Query Match Best Local Similarity 0.6%; Score 7; DB 1; Length 72;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1157 EEKREE 1163

DB 23 EEKREE 29

```

RESULT 42
DMS2_AGAN STANDARD; PRT; 73 AA.
AC 093222;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dermaseptin AA-2-5 precursor.
OS Agalychnis amae (Yellow-eye leaf frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Phyllomedusinae; Agalychnis.
OX NCBI_TaxID=75990;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Skin;
RX MEDLINE=98449786; PubMed=9774745;
RA Wechsclberger C.;
RT "Cloning of cDNAs encoding new peptides of the dermaseptin-family.";
RL Biochim. Biophys. Acta 1388:279-283(1998).
CC -1- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST GRAM-
CC POSITIVE AND GRAM-NEGATIVE BACTERIA. PROBABLY ACTS BY DISTURBING
CC MEMBRANE FUNCTIONS WITH ITS AMPHIPATHIC STRUCTURE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SKIN-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO THE DERMASEPTIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ005184; CAA06421.1; -
CC InterPro; IPR004275; Brevenin.
DR Pfam; PF03032; Brevenin; 1.
KW Antibiotic; Multigene family; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 42 POTENTIAL.
FT CHAIN 45 70 DERMASEPTIN AA-2-5.
FT PROPEP 72 73 POTENTIAL.
FT MOD_RES 70 70 AMIDATION (G-71 PROVIDE AMIDE GROUP)
SQ SEQUENCE 73 AA; 7855 MW; 4BDACDAD9DBF0944 CRC64;

```

Query Match Best Local Similarity 0.6%; Score 7; DB 1; Length 73;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1157 EEKREE 1163

```

Db          23 EEEKREE 29

RESULT 43
DMS3_PHYBI STANDARD; PRT; 74 AA.
ID_DMS3_PHYBI
AC P81485;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dermaseptin B11 precursor (Dermaseptin B3).
OS Phyllomedusa bicolor (Two-colored leaf frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Phyllomedusinae; Phyllomedusa.
OX NCBI_TaxId=8393;
RN [1]
SEQUENCE FROM N.A., SEQUENCE OF 46-73, AND MASS SPECTROMETRY.
RP TISSUE=Skin;
RX MEDLINE=98278974; PubMed=9614066;
RA Chapentier S., Aniche M., Mester J., Vouille V., Le Caer J.-P.,
RA Nicolas P., Delfour A., and molecular cloning of dermaseptins B, a
RT "Structure, synthesis, and molecular cloning of dermaseptins B, a
RT family of skin peptide antibiotics.";
RL J. Biol. Chem. 273:14690-14697(1998).
CC -1- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST GRAM-
CC POSITIVE AND GRAM-NEGATIVE BACTERIA. PROBABLY ACTS BY DISTURBING
CC MEMBRANE FUNCTIONS WITH ITS AMPHIPATHIC STRUCTURE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MASS SPECTROMETRY: MW=2780.4; MW ERR=0.1; METHOD=Electrospray.
CC -1- SIMILARITY: BELONGS TO THE DERMASEPTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: Y16564; CAA16288.1; -1-
CC DR InterPro: IPR004275; Brevinin.
CC DR Pfam: PF03032; Brevinin.1.
CC KW Antibiotic; Multigene family; Amphibian skin; Signal;
CC KV Cleavage on pair of basic residues.
CC FT SIGNAL 1 22 POTENTIAL.
CC FT PROPEP 23 43
CC FT CHAIN 46 74 DERMASEPTIN B11.
CC SQ SEQUENCE 74 AA; 8254 MW; F9D5902A24F32C8D CRC64;

Query Match 0.6%; Score 7; DB 1; Length 74;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1157 EEEKREE 1163
Db 23 EEEKREE 29

RESULT 44
ACP_COMTE STANDARD; PRT; 77 AA.
ID_ACP_COMTE
AC P80318;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acyl carrier protein (ACP).
GN ACP.
OS Comamonas testosteroni (Pseudomonas testosteroni).
OC Bacteria; Proteobacteria; Beta subdivision; Comamonadaceae; Comamonas.
OX NCBI_TaxId=285;
RN [1]

```

```

RP SEQUENCE.
RC STRAIN=ATCC 11996;
RX MEDLINE=97315245; PubMed=9171419;
RA Tang L., Weisenborn A.C., Kennedy E.P.;
RT "Domains of Escherichia coli acyl carrier protein important for
RT membrane-derived-oligosaccharide biosynthesis.";
RL J. Bacteriol. 179:3697-3705(1997).
CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty
CC acid biosynthesis.
CC -1- PATHWAY: De novo fatty acid biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- PTM: 4'-phosphopantetheine is transferred from CoA to a specific
CC serine of apo-ACP by acps. This modification is essential for
CC activity because fatty acids are bound in thioester linkage to the
CC sulfhydryl of the prosthetic group.
CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC HSP, P02901; IACP.
CC DR InterPro: IPR003231; Acyl_carrier.
CC DR InterPro: IPR003680; Pantine_attach.
CC DR Pfam: PF00550; pd-binding.1.
CC DR PRODOM: PD000887; Acyl_carrier.1.
CC DR TIGRFAMs: TIGR00517; acyl_carrier.1.
CC DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
CC DR PROSITE: PS00075; ACP_DOMAIN; 1.
CC KW Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine.
CC FT BINDING 36 36 PHOSPHOPANTHETHEINE.
CC SQ SEQUENCE 77 AA; 8406 MW; 2954BE8D07944B71 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 414 MALEDEF 420
Db 44 MALEDEF 50

RESULT 45
ACP1_RALSO STANDARD; PRT; 79 AA.
ID_ACP1_RALSO
AC O8Y0U1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acyl carrier protein 1 (ACP 1).
GN ACP1 OR ACP OR RSC1053 OR RS04175.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxId=305;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Broctier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigler P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty acid
CC biosynthesis (By similarity).
CC -1- PATHWAY: De novo fatty acid biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- PTM: 4'-phosphopantetheine is transferred from CoA to a specific
CC serine of apo-ACP by acps. This modification is essential for
CC activity because fatty acids are bound in thioester linkage to the
CC sulfhydryl of the prosthetic group (By similarity).
CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AL646062; CAD14755.1; -  
 DR InterPro: IPR003231; Acyl\_carrier.  
 DR InterPro: IPR003880; Ppantne\_attach.  
 DR Pfam: PF00550; pp-binding; 1.  
 DR ProDom: PD000887; Acyl\_carrier; 1.  
 DR TIGRFAMs: TIGR00517; acyl\_carrier; 1.  
 DR PROSITE: PS50075; ACP DOMAIN; 1.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.  
 KM Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;  
 KW Complete proteome.  
 FT BINDING 37 37 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 SQ SEQUENCE 79 AA; 8714 MW; F2737B0B5588E27B CRC64;

Query Match 0.6%; Score 7; DB 1; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 414 MALEDEF 420  
 Db 45 MALEDEF 51  
 |||||

RESULT 46  
 ACP\_XYLFA STANDARD; PRT; 79 AA.  
 ID ACP\_XYLFA  
 AC O9PFI5;  
 DT 15-JUN-2002 (Rel. 41, Last Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Acyl carrier protein (ACP).  
 GN ACP OR XP0672.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 CC Xylella.  
 NC NCBI\_TaxID=2371;  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.U.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 RA Barrios M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuriame E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silveira M.L.Z., Silveira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Trai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159(2000).  
 CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty acid  
 CC biosynthesis (By similarity).  
 CC -1- PATHWAY: De novo fatty acid biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- PTM: 4-phosphopantetheine is transferred from CoA to a specific  
 CC serine of apo-ACP by acps. This modification is essential for the  
 CC activity because fatty acids are bound in thioester linkage to the  
 CC sulfhydryl of the prosthetic group (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 ACRYL CARRIER DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE003911; AAF83482.1; ALT\_INIT.  
 DR HSSP: P02901; IACP.  
 DR InterPro: IPR003231; Acyl\_carrier.  
 DR InterPro: IPR003880; Ppantne\_attach.  
 DR Pfam: PF00550; pp-binding; 1.  
 DR ProDom: PD000887; Acyl\_carrier; 1.  
 DR TIGRFAMs: TIGR00517; acyl\_carrier; 1.  
 DR PROSITE: PS50075; ACP DOMAIN; 1.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.  
 KM Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;  
 KW Complete proteome.  
 FT BINDING 37 37 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 SQ SEQUENCE 79 AA; 8782 MW; 4A58822134BEC6B CRC64;

Query Match 0.6%; Score 7; DB 1; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 414 MALEDEF 420  
 Db 45 MALEDEF 51  
 |||||

RESULT 47  
 YAJC\_HAEIN STANDARD; PRT; 97 AA.  
 ID YAJC\_HAEIN  
 AC P44592;  
 DT 01-NOV-1995 (Rel. 32, Last Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DE Hypothetical protein HI0241.  
 GN HI0241.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Haemophilus.  
 NC NCBI\_TaxID=727;  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Keriavang A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitchugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Usterback T.R., Hama M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0092 FAMILY. STRONG. TO E.COLI YAJC.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U32710; AAC21909.1; -  
DR TIGR; H10241; -  
DR InterPro: IPR003849; Ya3C.  
DR Pfam: PF02699; DUF219; 1.  
DR TIGRFAMs; TIGR00739; Ya3C; 1.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRAASMEM 5 25 POTENTIAL.  
SQ SEQUENCE 97 AA; 10754 MW; 86F9A683A1D28358 CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 97;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1055 LAKGTEV 1061  
Db 42 LAKGTEV 48  
-----  
RESULT 48  
INS\_CHICK STANDARD; PRT; 107 AA.  
ID INS\_CHICK  
AC P01332;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Gallus gallus (Chicken),  
OS Meleagris gallopavo (Common turkey), and  
OS Struthio camelus (Ostrich).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031, 9103, 8801;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Chicken;  
RX MEDLINE=80222898; PubMed=7388949;  
RA Perler F., Efferatiadis A., Lomedico P., Gilbert W., Kolodner R.,  
RA Dodgson J.B.;  
RT "The evolution of genes: the chicken preproinsulin gene";  
RL Cell 20:555-566(1980).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Chicken; TISSUE=Pancreas;  
RX Hasegawa S., Honda K., Nata K., Yonekura H., Okamoto H., Hikami Y.,  
RT "Isolation of a cDNA encoding chicken insulin precursor";  
RL Anim. Sci. Technol. 62:867-869(1991).  
RN [3]  
RP SEQUENCE OF 25-54 AND 87-107.  
RC SPECIES=Chicken;  
RX MEDLINE=66160119; PubMed=5949593;  
RA Smith L.F.;  
RT "Species variation in the amino acid sequence of insulin";  
RL Am. J. Med. 40:662-666(1966).  
RN [4]  
RP SEQUENCE OF 25-54 AND 87-107.  
RC SPECIES=M.gallopavo;  
RX MEDLINE=72259992; PubMed=5066110;  
RA Jensch J.;  
RT "Structure and increased activity of insulin from the turkey  
RT (Meleagris gallopavo).";  
RL Hope-Seyler's Z. Physiol. Chem. 353:980-986(1972).  
RN [5]  
RP SEQUENCE OF 25-54 AND 87-107.

RC SPECIES=S.camelus;  
RX MEDLINE=88314456; PubMed=3045031;  
RA Evans T.K., Iltisauer D., Oelofsen W.;  
RT "Purification and primary structure of ostrich insulin";  
RL Int. J. Pept. Protein Res. 31:454-462(1988).  
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: THE PARTIAL SEQUENCE OF TURKEY AND OSTRICH ARE  
CC IDENTICAL WITH THAT OF CHICKEN.  
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J00874; AAA48925.1; -  
DR EMBL; J00872; AAA48925.1; JOINED.  
DR EMBL; V00416; CAA23707.1; -  
DR EMBL; V00418; CAA23708.1; -  
DR EMBL; X58993; CAA11738.1; -  
DR PIR; A01598; IPCR.  
DR PIR; A01599; INTR.  
DR PIR; JK0017; INOS.  
DR HSSP; P01308; IHS.  
DR InterPro: IPR004825; Ins/IGF/relax.  
DR Pfam: PF00049; Insulin.1.  
DR PRINTS; PR00276; INSULIN.1.  
DR PRINTS; PR00277; INSULIN.1.  
DR SMART; SM00078; IIGF.1.  
DR PROSITE; PS00262; INSULIN; 1.  
KW Insulin family; Hormone; Glucose metabolism; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 54 INSULIN B CHAIN.  
FT PROPEP 57 84 C PEPTIDE.  
FT CHAIN 87 107 INSULIN A CHAIN.  
FT DISULFID 31 93 INTERCHAIN.  
FT DISULFID 43 106 INTERCHAIN.  
FT DISULFID 92 97.  
SQ SEQUENCE 107 AA; 11981 MW; 3D43C5D82B25DCDD CRC64;  
  
Query Match 0.6%; Score 7; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1136 LIALIVF 1142  
Db 10 LIALIVF 16  
-----  
RESULT 49  
HDEA\_ECOLI STANDARD; PRT; 110 AA.  
ID HDEA\_ECOLI  
AC P26604;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Protein hdea precursor (10k-S protein).  
GN HDEA OR B3510 OR Z4922 OR ECS4390.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-40.  
 RC STRAIN=K12;  
 RX MEDLINE=93204884; PubMed=8455549;  
 RA Yoshida T., Ueguchi C., Yamada H., Mizuno T.;  
 RT "Function of the Escherichia coli nucleoid protein, H-NS: molecular  
 RT analysis of a subset of proteins whose expression is enhanced in a  
 RT hns deletion mutant.";  
 RT Mol. Gen. Genet. 237:113-122(1993).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes.";  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G.J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobo T.,  
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RN [15]  
 RP GENE NAME.  
 RX MEDLINE=94064579; PubMed=8244952;  
 RA Yoshida T., Ueguchi C., Mizuno T.;  
 RT "Physical map location of a set of Escherichia coli genes (hde) whose  
 RT expression is affected by the nucleoid protein H-NS.";  
 RL J. Bacteriol. 175:7747-7748(1993).  
 RN [16]  
 RP SEQUENCE OF 22-41.  
 RC STRAIN=K12 / W3110;  
 RX Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,  
 RA Fruiter S., Paquet N., Wilkins M., Appel R.D., Baloch A.,  
 RA Hochstrasser D.F.;  
 RT Submitted (SEP-1994) to the SWISS-PROT data bank.  
 RN [17]  
 RP SEQUENCE OF 22-33.  
 RC STRAIN=K12 / EMG2;  
 RX MEDLINE=97443975; PubMed=9298646;  
 RA Link A.J., Robison K., Church G.M.;  
 RT "Comparing the predicted and observed properties of proteins encoded  
 RT in the genome of Escherichia coli K-12.";  
 RL Electrophoresis 18:1259-1313(1997).  
 RN [18]  
 RP SEQUENCE OF 22-31.  
 RC STRAIN=K12;  
 RX MEDLINE=99065675; PubMed=9868784;  
 RA Wasinger V.C., Humphery-Smith I.;  
 RT "Small genes/gene-products in Escherichia coli K-12.";  
 RL FEMS Microbiol. Lett. 169:375-382(1998).  
 RN [19]  
 RP X-RAY CRYSTALLOGRAPHY  
 RX MEDLINE=98400493; PubMed=9731767;  
 RA Yang F., Gustafson K.R., Boyd M.R., Wlodawer A.;  
 RT "Crystal structure of Escherichia coli HdeA.";

RL Nat. Struct. Biol. 5:763-764(1998).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=20090957; PubMed=10623550;  
 RA Gajiwala K.S., Burley S.K.;  
 RT "HdeA, a periplasmic protein that supports acid resistance in  
 RT pathogenic enteric bacteria.";  
 RL J. Mol. Biol. 295:605-612(2000).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL, D11109; BAA01883.1; -;  
 CC EMBL, U00039; AAB18486.1; -;  
 CC EMBL, AE000427; AAC76535.1; -;  
 CC EMBL, AE005576; AAG58651.1; -;  
 CC EMBL, AF002565; BAB37813.1; -;  
 CC PIR, S30266; S30266.  
 CC PIR, S30268; S30268.  
 CC PDB, 1BG8; 16-SEP-98.  
 CC PDB, 1D78; 10-DEC-99.  
 CC SWISS-2DPAGE; P26604; COLI.  
 CC Ecogene; Ecol198; hdeA.  
 CC Periplasmic; Signal; 3D-structure; Complete proteome.  
 FT SIGNAL 1 21  
 FT CHAIN 22 110 PROTEIN HDEA.  
 FT DISULFID 39 87  
 SQ SEQUENCE 110 AA; 11858 MW; 063262C4863FA2E9 CRC64;  
 QY 1131 IGGILLL 1137  
 Db 9 IGGILLL 15  
 RESULT 50  
 IGF MYXGL  
 ID IGF MYXGL STANDARD; PRT; 139 AA.  
 AC P22618;  
 DT 01-MAY-1991 (Rel. 18; Created)  
 DT 01-MAY-1991 (Rel. 18; Last sequence update)  
 DT 01-FEB-1994 (Rel. 28; Last annotation update)  
 DE Insulin-like growth factor precursor (IGF) (Fragment).  
 OS Myxine glutinosa (Atlantic hagfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;  
 OC Myxiniidae; Myxiniinae; Myxine.  
 OC NCBI\_Taxid=7769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91115860; PubMed=1989990;  
 RA Nagamatsu S., Chan S.J., Falkner S., Steiner D.F.;  
 RT "Evolution of the insulin gene superfamily. Sequence of a  
 RT preproinsulin-like growth factor cDNA from the Atlantic hagfish.";  
 RL J. Biol. Chem. 266:2397-2402(1991).  
 CC -1- FUNCTION: THE INSULIN-LIKE GROWTH FACTORS, ISOLATED FROM PLASMA,  
 CC ARE STRUCTURALLY AND FUNCTIONALLY RELATED TO INSULIN BUT HAVE A  
 CC MUCH HIGHER GROWTH-PROMOTING ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; M57735; AAA49265.1; -  
 DR PIR; A38612; A38612.  
 DR HSSP; P01344; 1GF2.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KM Insulin family; Mitogen; Growth factor; Signal.  
 FT NON\_TER 1  
 FT SIGNAL <1 38  
 FT CHAIN 39 139 INSULIN-LIKE GROWTH FACTOR.  
 FT DOMAIN 39 67 B.  
 FT DOMAIN 68 82 C.  
 FT DOMAIN 83 103 A.  
 FT DOMAIN 104 113 D.  
 FT DOMAIN 114 139 E.  
 SQ SEQUENCE 139 AA; 16087 MW; 2FC88C8D074FAC1 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLALL 1140  
 DB 28 LLLALL 34

RESULT 51  
 CDD\_YEAST STANDARD; PRT; 142 AA.  
 AC 006549;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytidine deaminase (EC 3.5.4.5) (Cytidine aminohydrolase) (CDA).  
 GN CDD1 OR YLR245C OR I9672.13.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=FL100;  
 RA MEDLINE=99431800; PubMed=10501935;  
 RA Kurtz J.-E., Exinger F., Eids P., Jund R.;  
 RA "New insights into the pyrimidine salvage pathway of Saccharomyces  
 RT cerevisiae: requirement of six genes for cytidine metabolism.";   
 RL Curr. Genet. 36:130-136(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288c / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Pavello A., Patton L., Gattung S., Greco T., Kirsten J.,  
 RA Kucaba T., Hallsworth K., Hawkins J., Latreille P., Le T.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Martin E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Riles L., Riles L., Taich A., Trevasakis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterson R.;  
 RA Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: THIS ENZYME SCAVENGE EXOGENOUS AND ENDOGENOUS CYTIDINE  
 AND 2'-DEOXYCYTIDINE FOR UMP SYNTHESIS.  
 CC - CATALYTIC ACTIVITY: Cytidine + H(2)O = uridine + NH(3).  
 CC - COFACTOR: ZINC (BY SIMILARITY).  
 CC - SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES  
 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; AF080089; AAD04031.1; -  
 DR EMBL; U20865; AAB67399.1; -  
 DR HSSP; P13652; 1ALN.  
 DR SGD; S0004235; CDD1.  
 DR InterPro; IPR002125; dCMP/cyt deam.  
 DR Pfam; PF00383; dCMP\_cyt\_deam; 1.  
 DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; 1.  
 KM Hydroxylase; Zinc.  
 FT METAL 61 61 ZINC (BY SIMILARITY).  
 FT METAL 96 96 ZINC (BY SIMILARITY).  
 FT METAL 99 99 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 142 AA; 15536 MW; 962BD9CCB30D51F CRC64;

Query Match 0.6%; Score 7; DB 1; Length 142;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 287 AALKACE 293  
 DB 17 AALKACE 23

RESULT 52  
 MK\_CHICK STANDARD; PRT; 142 AA.  
 ID MK\_CHICK  
 AC P24052;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Midkine precursor (retinoic acid-induced heparin-binding protein)  
 DE (R1-HB).  
 GN RHIB.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 80-85; 112-118 AND 121-133.  
 RC TISSUE=Embryo;  
 RA MEDLINE=91207359; PubMed=2018506;  
 RA Urios P., Duprez D., le Caer J.-P., Coutrois Y., Vigny M., Laurent M.;  
 RA "Molecular cloning of R1-HB, a heparin binding protein regulated by  
 RT retinoic acid.";   
 RL Biochem. Biophys. Res. Commun. 175:617-624(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=95010085; PubMed=7925417;  
 RA Duprez D., Treaggar J., Pecqueur C., Vigny M.R.;  
 RA "Organisation and promoter activity of the  
 RT retinoic-acid-induced-heparin-binding (RHIB) gene.";   
 RL Eur. J. Biochem. 224:931-941(1994).  
 RN [3]  
 RP SEQUENCE OF 22-77.  
 RA MEDLINE=91128406; PubMed=1993066;  
 RA Raulais D., Lagente-Chevallier O., Guetret C., Duprez D., Coutrois Y.,  
 RA Vigny M.;  
 RA "A new heparin binding protein regulated by retinoic acid from chick  
 RT embryo.";   
 RL Biochem. Biophys. Res. Commun. 174:708-715(1991).  
 RN [4]  
 RP SEQUENCE OF 22-77.  
 RC STRAIN=White leghorn; TISSUE=Embryo;  
 RX MEDLINE=90108010; PubMed=2558016;  
 RA Vigny M., Raulais D., Puzenat N., Duprez D., Hartman M.P.,  
 RA Jeanmy J.C., Coutrois Y.;  
 RA "Identification of a new heparin-binding protein localized within  
 RT chick basement membranes.";

```

RL Bur. J. Biochem. 186:733-740(1989).
CC -1- FUNCTION: HAS MITOGENIC ACTIVITY, AND NEURITE EXTENSION ACTIVITY
CC FOR PC12 CELLS.
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES IN EARLY EMBRYONIC
CC TISSUES, AND CELL SURFACE OF NEUROECTODERMAL CELLS.
CC -1- DEVELOPMENTAL STAGE: ESSENTIALLY EXPRESSED DURING EMBRYOGENESIS.
CC -1- INDUCTION: BY RETINOIC ACID.
CC -1- SIMILARITY: BELONGS TO THE PLEIOTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M61754; -; NOT ANNOTATED; CDS.
CC PIR; J76482; CAA54020.1; -.
CC HSSP; P21741; 1MKC.
CC InterPro; IPR000762; PTN_MK.
CC Pfam; PFO1091; PTN_MK; 1.
CC PRINTS; PR00269; PTNMDKNE.
CC ProDom; PD005592; PTN_MK; 1.
CC SMART; SM00193; PTN; 1.
CC PROSITE; PS00619; PTN_MK_1; 1.
CC PROSITE; PS00620; PTN_MK_2; 1.
CC Growth factor; Mitogen; Differentiation; Heparin-binding; Signal.
CC FT CHAIN 1 21
CC FT DISULFID 22 142 MDKNE.
CC FT DISULFID 36 60 BY SIMILARITY.
CC FT DISULFID 44 69 BY SIMILARITY.
CC FT DISULFID 51 73 BY SIMILARITY.
CC FT DISULFID 83 115 BY SIMILARITY.
CC FT DISULFID 93 125 BY SIMILARITY.
CC FT CONFLICT 89 89 S -> R (IN REF. 2).
CC FT CONFLICT 93 93 C -> G (IN REF. 2).
CC FT CONFLICT 93 93
CC SQ SEQUENCE 142 AA; 15579 MW; 9D05CAF9558451B CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 142;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLLLL 1140
DB 7 LLLLLL 13

RESULT 53
GLB_APLJU STANDARD; PRT; 144 AA.
AC P14393;
ID GLB_APLJU
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Globin (Myoglobin).
OS Aplysia juliana (Sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspeidae;
OC Aplysiidae; Aplysia.
OC NCB1_TaxID=6506;
OX NCB1_TaxID=6506;
RN (1)
RX MEDLINE=85134891; PubMed=6527386;
RT Takagi T., Iida S., Matsuo K., Shikama K.;
RT "Aplysia myoglobins with an unusual amino acid sequence.";
RL J. Mol. Biol. 180:1179-1184(1984).
CC -1- SUBUNIT: Monomer.
CC -1- MISCELLANEOUS: THIS MOLUSCAN GLOBIN LACKS ONE OF THE HEME-BINDING
CC HISTIDINE RESIDUES FOUND IN MOST OTHER GLOBINS.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC PIR; A44585; A44585.
CC HSSP; P02210; 2FAL.

```

```

DR InterPro; IPR002336; Erythrorin.
DR InterPro; IPR000971; Globin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00611; ERYTHRUCORIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle; Acetylation.
FT MOD RES 1 1 ACETYLATION.
FT METAL 95 95 IRON (HEME PROXIMAL LIGAND)
FT METAL 95 95 (BY SIMILARITY).
SQ SEQUENCE 144 AA; 15016 MW; 28FCF0FC578E50FB CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 144;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 984 GLIISAL 990
DB 133 GLIISAL 139

RESULT 54
RHC_RHIV
ID RHC_RHIV STANDARD; PRT; 149 AA.
AC Q03315;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein rhic precursor.
DE RHC.
GN RHC.
OS Rhizobium leguminosarum (biovar viciae).
OG Plasmid sym pRL101.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCB1_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92283758; PubMed=1597418;
RA Cubo M.T., Economou A., Murphy G.J., Johnston A.W., Downie J.A.;
RT "Molecular characterization and regulation of the
RT rhizosphere-expressed genes rhiaBCR that can influence nodulation by
RT Rhizobium leguminosarum biovar viciae.";
RL J. Bacteriol. 174:4026-4035(1992).
CC -1- FUNCTION: MAY BE INVOLVED IN PLANT-MICROBE INTERACTION.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M98835; AAA26359.1; -.
CC PIR; C41887; C41887.
KW Plasmid; Periplasmic; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 1 149 PROTEIN RHC.
FT CHAIN 24 149
SQ SEQUENCE 149 AA; 15299 MW; F41A72AB1646B193 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 149;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 754 ALTVTFA 760
DB 15 ALTVTFA 21

RESULT 55
BIR_HUMAN STANDARD; PRT; 160 AA.
ID BIR_HUMAN
AC Q13523; Q16582;

```

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Bcl-2 interacting killer (Apoptosis inducer NBK) (BP4) (BIP1).  
 GN BIK OR NBK.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 ON NCBI TaxID=9606;  
 RX MEDLINE=96060622; PubMed=7478623;  
 RA Boyd J.M., Gallo G.J., Elangovan B., Houghton A.B., Malstrom S.,  
 RA Avery B.J., Ebb R.G., Subramanian T., Chittenden T., Lutz R.J.,  
 RA Chinnadurai G.;  
 RT "Bik, a novel death-inducing protein shares a distinct sequence motif  
 RT with Bcl-2 family proteins and interacts with viral and cellular  
 RT survival-promoting proteins.";  
 RT Oncogene 11:1921-1928(1995).  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=B-cell;  
 RX MEDLINE=96060622; PubMed=7478623;  
 RA Boyd J.M., Gallo G.J., Elangovan B., Houghton A.B., Malstrom S.,  
 RA Avery B.J., Ebb R.G., Subramanian T., Chittenden T., Lutz R.J.,  
 RA Chinnadurai G.;  
 RT "Bik, a novel death-inducing protein shares a distinct sequence motif  
 RT with Bcl-2 family proteins and interacts with viral and cellular  
 RT survival-promoting proteins.";  
 RT Oncogene 11:1921-1928(1995).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96413338; PubMed=8816500;  
 RA Han J., Sabbatini P., White E.;  
 RT "Induction of apoptosis by human NBK/Bik, a Bcl-2-containing protein  
 RT that interacts with Bcl-2.";  
 RT Mol. Cell. Biol. 16:5857-5864(1996).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymphoid;  
 RA Pun K.-T., Farrow S.N., Raven T., Wride C.J., White J.H.M., Brown R.;  
 RT Submitted (Jun-1995) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99431905; PubMed=10500065;  
 RA Casellas A., Ino Y., Louis D.N., Ramesh V., Gusella J.F., Rustgi A.K.;  
 RT "Mapping of a target region of allelic loss to a 0.5-cM interval on  
 RT chromosome 22q13 in human colorectal cancer.";  
 RT Gastroenterology 117:831-837(1999).  
 RN (5)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Buskiewicz R., Beare D.M.,  
 RA Clump M., Smith L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Cordy N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall S.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
 RA McElay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.C.T.,  
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
 RA Scott C.E., Senter H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soudland C., Spurgeon L., Stewart C.A., Sultson J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshitaki Y., Aoki N., Mitsuyma S.,  
 RA Roe B.A., Chen F., Chu Y., Hu P., Hua A., Kenton S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Do A., Do T.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren O., Salaj E., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
 RA Zhang W., Zhang G., Chisoe S., Murray J., Miller N., Wink P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,

RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latselle P., Layman D., Ozerky P., Rohlfing T.,  
 RA Schick P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
 RA Korfi I., Bedell J.A., Hillier L., Mardis E., Weststrom R., Wilson R.,  
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saita S., Budarf M.L.,  
 RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 RA Kim U.J., Shizuya H., Simon M.I., Dunamski J.P., Peyrard M., Kedra D.,  
 RA Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,  
 RA Wilkerson P., Bodencich A., Hartman K., Hu X., Khan A.S., Lane L.,  
 RA Tilahun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RT Nature 402:489-495(1999).  
 RN (6)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph;  
 RA Strausberg R.;  
 RT Submitted (Jan-2001) to the EMBL/GenBank/DBJ databases.  
 RN (7)  
 RP MUTAGENESIS, AND FUNCTION OF BCL-2 DOMAIN.  
 RX MEDLINE=96091131; PubMed=8521816;  
 RA Chittenden T., Flemington C., Houghton A.B., Ebb R.G., Gallo G.J.,  
 RA Elangovan B., Chinnadurai G., Lutz R.J.;  
 RT "A conserved domain in Bak, distinct from Bcl-1 and Bcl-2, mediates cell  
 RT death and protein binding functions.";  
 RT EMBO J. 14:5589-5596(1995).  
 CC -1- FUNCTION: ACCELERATES PROGRAMMED CELL DEATH. BINDING TO THE  
 CC APOPTOSIS REPRESSORS BCL-X(L), BHRF1, BCL-2 OR ITS ANOGENOUS  
 CC HOMOLOG BCL-19K PROTEIN SUPPRESSES THIS DEATH-PROMOTING ACTIVITY.  
 CC DOES NOT INTERACT WITH BAX.  
 CC -1- SUBCELLULAR LOCATION: AROUND THE NUCLEAR ENVELOPE. AND IN  
 CC CYTOPLASMIC MEMBRANES.  
 CC -1- DOMAIN: INTACT BCL-2 DOMAIN IS REQUIRED BY BIK, BID, BAK, BAD AND  
 CC BAX FOR THEIR PRO-APOPTOTIC ACTIVITY AND FOR THEIR INTERACTION  
 CC WITH ANTI-APOPTOTIC MEMBERS OF THE BCL-2 FAMILY.  
 CC APOPTOTIC MEMBERS OF THE BCL-2 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG 3 (BCL-2) DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, U34584; AAC50413.1; -;  
 DR EMBL, U49730; AAC79124.1; -;  
 DR EMBL, X89986; CA62013.1; -;  
 DR EMBL, AF174424; AAF01156.1; -;  
 DR EMBL, AF174421; AAF01156.1; JOINED.  
 DR EMBL, AF174422; AAF01156.1; JOINED.  
 DR EMBL, AF174423; AAF01156.1; JOINED.  
 DR EMBL, AL022237; CA18260.2; -;  
 DR EMBL, BC001599; AA01599.1; -;  
 DR GeneW; HGNC:1051; BIK.  
 DR MIM: 603392; -;  
 DR InterPro; IPR000712; Bcl2\_BH.  
 DR PROSITE; PS01259; BH3; 1.  
 KW Apoptosis; Transmembrane.  
 FT DOMAIN 57 71 BH3.  
 FT TRANSMEM 136 156 POTENTIAL.  
 FT DOMAIN 137 158 LEUCINE-ZIPPER (POTENTIAL).  
 FT CONFLICT 149 150 PL->LP (IN REF. 1).  
 SQ SEQUENCE 160 AA; 18016 MW; 890344443F5A136 CRC64;  
 QY 1134 LLLALL 1140  
 DB 141 LLLALL 147  
 Query Match 0.6%; Score 7; DB 1; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 56
YCRS MYXXA      STANDARD;      PRT;      170 AA.
ID YCRS MYXXA
AC 006929;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 18.1 kDa protein in CARS 3'region (ORF4).
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cytoobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DK101;
RX MEDLINE=95020544; PubMed=7934835;
RA McGowan S.D., Gorham H.C., Hodgson D.A.;
RT "Light-induced carotenogenesis in Myxococcus xanthus: DNA sequence
RT analysis of the car region.";
RL Mol. Microbiol. 10:713-735(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X71062; CAA50384.1; -
DR PIR; S33152; S33152.
DR PIR; S39880; S39880.
DR PIR; S35742; S35742.
KM Hypothetical protein.
SQ SEQUENCE 170 AA; 18052 MW; E446D38979EBF02D CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 170;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 403 LEGHRL 409
DB 22 LEGHRL 28

RESULT 57
ATPF SYNPE      STANDARD;      PRT;      171 AA.
ID ATPF SYNPE
AC P08447;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase B chain (EC 3.6.3.14) (Subunit I).
GN ATPF.
OS Synecococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=1139;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=87311713; PubMed=3041005;
RA Cozens A.L., Walker J.E.;
RT "The organization and sequence of the genes for ATP synthase subunits
RT in the cyanobacterium Synecococcus 6301. Support for an
RT endosymbiotic origin of chloroplasts.";
RL J. Mol. Biol. 194:359-383(1987).
RL J. Mol. Biol. 194:359-383(1987).
CC -----
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -----
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -----
CC -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05102; CAA28926.1; -
DR PIR; S10829; LMVCL.
DR InterPro; IPR002146; ATPsynt_B/B' sub.
DR Pfam; PF00430; ATP-synt_B; 1.
KM Hydrogen ion transport; Transmembrane; CF(0).
FT TRANSMEM 24 40
FT POTENTIAL.
SQ SEQUENCE 171 AA; 18681 MW; ADB85FE8F4D98C2C CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 171;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 EARLLV 272
DB 86 EARLLV 92

RESULT 58
YF87 METJA      STANDARD;      PRT;      171 AA.
ID YF87 METJA
AC Q58982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1587.
GN MJ1587.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodex A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Furumann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
RL Science 273:1058-1073(1996).
CC -----
CC -1- SIMILARITY: TO M.JANNASCHII MJ0417.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67899; AAB99616.1; -
DR TIGR; MJ1587; -
DR InterPro; IPR004013; PNP_C.
DR InterPro; IPR003141; PNP_N.
DR Pfam; PF02231; PNP_N; 1.
DR Pfam; PF02811; PNP_C; 1.
DR SMART; SM00481; POLI11AC; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 171 AA; 19198 MW; 7D4F7B936C5C7472 CRC64;

```

Query Match 0.6%; Score 7; DB 1; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 ALGDRIF 355  
 |||||  
 DB 123 ALGDRIF 129

RESULT 59  
 PORC\_METUA STANDARD; PRT; 178 AA.  
 ID 057717;  
 AC 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pyruvate synthase subunit PORC (EC 1.2.7.1). (Pyruvate oxidoreductase  
 gamma chain) (POR) (Pyruvic-ferredoxin oxidoreductase gamma subunit).  
 GN PORC OR Mj0269.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 CC Methanocaldococcaceae; Methanocaldococcus.  
 OK NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Keriakova A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Ullrich T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Corton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073 (1996).  
 CC -1- CATALYTIC ACTIVITY: Pyruvate + CoA + oxidized ferredoxin = acetyl-  
 -1- SUBUNIT: HETEROLEPTIMER OF ONE ALPHA, ONE BETA, ONE DELTA AND ONE  
 -1- GAMMA CHAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U67482; AAB88256.1; -  
 CC TIGR: Mj0269; -  
 CC DR InterPro: IPR002869; POR.  
 CC DR Pfam: PF01558; POR; 1.  
 CC KW Oxidoreductase; Complete proteome.  
 CC SQ SEQUENCE 178 AA; 19200 MW; 3758E7E7C70FAEEL CRC64;

Query Match 0.6%; Score 7; DB 1; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 LKKGAV 471  
 |||||  
 DB 87 LKKGAV 93

RESULT 60  
 HYFH\_ECOLI STANDARD; PRT; 181 AA.  
 ID HYFH\_ECOLI  
 AC P77423;  
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hydrogenase-4 component H.  
 GN HYFH OR B2488.  
 OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OK NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Andrews S.C., McClay J., Ambler A., Quail M., Berts B.C., Guest J.R.,  
 RA Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1233-1244 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97349980; PubMed=9205837;  
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kikagawa M., Makino K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubraman S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horuchi T.;  
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113 (1997).  
 CC -1- FUNCTION: PROBABLE ELECTRON TRANSFER PROTEIN FOR HYDROGENASE 4.  
 CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF  
 CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M61654; AAB88570.1; -  
 CC DR EMBL: AE000335; AAC75541.1; -  
 CC DR EMBL: D90877; BAA16376.1; -  
 CC DR HSRP; P00195; ICLP.  
 CC DR Ecogene; EG14216; hyfH.  
 CC DR InterPro: IPR001450; 4Fe4S\_ferredoxin.  
 CC DR Pfam: PF00037; Fe4S; 2.  
 CC DR PROSITE: PS00198; 4FE4S\_FERREDOXIN; 2.  
 CC KW Electron transport; 4Fe-4S; Iron-sulfur; Complete proteome.

FT METAL 40 40 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
 FT METAL 43 43 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
 FT METAL 46 46 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
 FT METAL 50 50 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 75 75 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 78 78 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 81 81 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 85 85 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 FT METAL 85 85 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 181 AA; 20159 MW; 5DBBA3C7C7EBD885 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 181;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 TDVLVA 507

Db 171 TDVLVA 177

RESULT 61  
UVP1\_ECOLI STANDARD; PRT; 198 AA.

AC P18957;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)

GN UVP1 protein.

OS Escherichia coli.

OC Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Escherichia

OX NCBI\_TaxId=562;

RN (1)

RP SEQUENCE FROM N.A.

RA MEDLINE=89384434; PubMed=2550763;

RG Gigliani F., Sporeno E., Perri S., Battaglia P.A.;

RT "The uvp1 gene of plasmid pr cooperates with mucAB genes in the DNA

repair process.";

RL Mol. Gen. Genet. 218:18-24(1989).

CC -1- FUNCTION: COOPERATES WITH MUCAB GENES IN THE DNA REPAIR PROCESS.

CC IT COULD BE A RESOLVASE-INVERTASE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE SITE-SPECIFIC RECOMBINASE RESOLVASE

FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X16119; CAA34249.1; -

DR PIR; S04903; S04903.

DR HSSP; P03012; 2RSL.

DR InterPro; IPR001822; Recombinase.

DR Pfam; PF00239; HTH\_7; 1.

DR Pfam; PF00239; HTH\_7; 1.

DR PROSITE; PS00397; RECOMBINASES\_1; FALSE\_NEG.

DR PROSITE; PS00398; RECOMBINASES\_2; 1.

KW DNA repair: plasmid; DNA-binding; DNA recombination; DNA integration;

KM DNA invertase.

FT ACT\_SITE 9

FT TRANSIENT COVALENT LINKAGE TO DNA DURING

STRAND CLEAVAGE AND REJOINING

(BY SIMILARITY).

FT DNA BIND 168 187 H-T-H MOTIF (BY SIMILARITY).

FT SEQUENCE 198 AA; 22127 MW; 8011E768EB0548 CRC64;

SO

Query Match 0.6%; Score 7; DB 1; Length 198;

Best Local Similarity 100.0%; Pred. No. 74;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1003 SUSQVIT 1009

Db 74 SUSQVIT 80

RESULT 62

RACI\_DICDI STANDARD; PRT; 205 AA.

AC Q9GPR2;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE PAS-related protein rac1.

GN RAC1.

OS Dictyostelium discoideum (slime mold).

OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

OX NCBI\_TaxId=44689;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=AX4;

RA MEDLINE=21127961; PubMed=11222756;

RT Rivero F., Dislich H., Glockner G., Noegel A.A.;

RL "The Dictyostelium discoideum family of Rho-related proteins.";

CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF310895; AAG45138.1; -

DR HSSP; P15154; 1B96.

DR DictyDb; DD27272; rac1.

DR InterPro; IPR003578; GTPase\_Rho.

DR InterPro; IPR001230; Prenyl\_site.

DR InterPro; IPR001806; Ras\_trnstrng.

DR Pfam; PF00071; ras; 1.

DR PRINTS; PR00449; RASFRNFRMNG.

DR SMART; SM00174; RHO; 1.

DR TIGRfams; TIGR00231; small\_GTP\_1.

KW GTP-binding; Prenylation; Lipoprotein.

FT NP BIND 12 19 GTP (BY SIMILARITY).

FT NP BIND 59 63 GTP (BY SIMILARITY).

FT NP BIND 119 122 GTP (BY SIMILARITY).

FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).

FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).

FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).

FT SEQUENCE 205 AA; 22997 MW; F9EFD31576C45CC CRC64;

SO

Query Match 0.6%; Score 7; DB 1; Length 205;

Best Local Similarity 100.0%; Pred. No. 76;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1046 QVVRCHL 1052

Db 176 QVVRCHL 182

RESULT 63

GPBB\_MOUSE STANDARD; PRT; 206 AA.

AC P56400;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Platelet glycoprotein Ib beta chain precursor (GP-Ib beta) (GP1bB)

DE (GP1b-Beta).

GN GP1B8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxId=10090;

RN (1)

RP SEQUENCE FROM N.A.

RA MEDLINE=97403789; PubMed=9259114;

RT Kiteguchi T., Murata M., Anbo H., Moriki T., Ikeda Y.;

RL "Characterization of the gene encoding mouse platelet glycoprotein Ib

beta.";

RT Thromb. Res. 87:235-244(1997).

CC -1- FUNCTION: GP-IB, A SURFACE MEMBRANE PROTEIN OF PLATELETS,

PARTICIPATES IN THE FORMATION OF PLATELET PLUGS BY BINDING TO VON

WILLEBRAND FACTOR, WHICH IS ALREADY BOUND TO THE SUBENDOTHELIUM

(BY SIMILARITY).

CC -1- SUBUNIT: GP-IB ALPHA AND BETA ARE DISULFIDE LINKED. GP-IX IS

```

CC COMPLEXED WITH THE GP-IB HETERODIMER VIA A NON COVALENT LINKAGE
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: PLATELET ACTIVATION APPARENTLY INVOLVES DISRUPTION
CC OF THE MACROMOLECULAR COMPLEX OF GP-IB WITH THE PLATELET
CC GLYCOPROTEIN IX (GP-IX) AND DISSOCIATION OF GP-IB FROM THE
CC ACTIN-BINDING PROTEIN.
CC -1- SIMILARITY: CONTAINS 1 LEUCINE-RICH REPEAT (LRR).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AB001419; BAA22424.1; -.
CC DR MGD; MGI:107852; GpIb.
CC DR InterPro; IPR001611; LRR.
CC DR InterPro; IPR000483; LRR_Cterm.
CC DR InterPro; IPR000372; LRR_Nterm.
CC DR Pfam; PRO0560; LRR_1.
CC DR Pfam; PF01462; LRRNT; 1.
CC DR Pfam; PF01463; LRRCT; 1.
CC DR SMART; SM00082; LRRCT; 1.
CC DR SMART; SM00013; LRRNT; 1.
CC DR Platelet; Transmembrane; Glycoprotein; Hemostasis; Blood coagulation;
CC Signal; Phosphorylation; Cell adhesion; Leucine-rich repeat.
CC KW SIGNAL
CC FT CHAIN 1 26
CC FT DOMAIN 27 206
CC FT TRASMEM 148 172
CC FT DOMAIN 173 206
CC FT REPEAT 60 83
CC FT CAROXYD 66
CC FT MOD RES 191
CC FT SEQUENCE 206 AA; 21762 MW; AC4BCB4DFA226F1D CRC64;
SQ
Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 206;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1134 LLLALL 1140
Db 13 LLLALL 19

```

```

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean C.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymptre B.,
RA Weljens I., Vanstreels E., Rieger M., Scheer M., Mueller-Auer S.,
RA Gabel K., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambolt R., Punnett B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AB016216; BAA1740.1; -.
CC DR EMBL; AL136235; CAB65806.1; -.
CC DR InterPro; IPR001380; Ribosomal_L13E.
CC DR Pfam; PF01294; Ribosomal_L13e; 1.
CC DR Prodom; PD004443; Ribosomal_L13E; 1.
CC DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
CC KW Ribosomal protein.
CC SEQUENCE 208 AA; 23529 MW; B98EF67308F53F45 CRC64;
SQ
Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 208;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 709 AGARAF 715
Db 189 AGARAF 195

```



RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Suleon J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES  
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND  
 CC THEN IS RELEASED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; 277165; CAB01009.1; -  
 DR EMBL; AE007157; AAK47861.1; ALT\_INIT.  
 DR TIGR; MT3523; -  
 CC -----  
 DR Tuberculosis; Rv3414c; -  
 DR InterPro; IPR000838; Sigma70\_ECF.  
 DR Pfam; PF00776; Sigma70\_ECF; 1.  
 DR PROSITE; PS01063; SIGMA70\_ECF; 1.  
 KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;  
 KW DNA-binding; Complete proteome.  
 FT DOMAIN 75 88  
 FT DNA\_BIND 176 195 H-T-H MOTIF (BY SIMILARITY).  
 FT SEQUENCE 212 AA; 22919 MW; 12A157F6F66B0C3 CRC64;  
 SQ  
 Query Match 0.6%; Score 7; DB 1; Length 212;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 469 GAVRAQ 475  
 Db 188 GAVRAQ 194  
 RESULT 66  
 ID 10.0 CAEEL STANDARD; PRT; 214 AA.  
 AC 009533;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 60S ribosomal protein L10 (QM protein homolog).  
 GN RPL-10 OR F10B5.1.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Poliderrinae; Caenorhabditis.  
 CC NCB1\_TaxID=62319;  
 OK NCB1\_TaxID=62319;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Simms M.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE L10E FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; 248334; CAA8308.1; -  
 DR WormPep; F10B5.1; CE01543.  
 DR InterPro; IPR001197; Ribosomal\_L10E.  
 DR Pfam; PF00826; Ribosomal\_L10E; 1.  
 DR TIGRFAMs; TIGR00279; L10E; 1.  
 DR PROSITE; PS01257; RIBOSOMAL\_L10E; 1.  
 KW Ribosomal protein.  
 FT SIGNAL 1 19  
 FT CHAIN 20 220  
 FT CARBOHYD 66 66 N-LINKED (GLCNAC...); (POTENTIAL).  
 FT SEQUENCE 220 AA; 25050 MW; B15D41D1551C5D5 CRC64;  
 SQ  
 Query Match 0.6%; Score 7; DB 1; Length 214;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1074 FRRAKK 1080  
 Db 152 FRRAKK 158  
 RESULT 67  
 ID FGF3 CHICK STANDARD; PRT; 220 AA.  
 AC P48801;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Fibroblast growth factor-3 precursor (FGF-3) (HBGF-3).  
 GN FGF3 OR FGF-3. (Chicken).  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 OK NCB1\_TaxID=9031;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rhode Island red; TISSUE=Embryo;  
 RX MEDLINE=95309122; PubMed=7789270;  
 RA Mahmood R., Kiefer P., Guthrie S., Dickson C., Mason I.;  
 RT "Multiple roles for FGF-3 during cranial neural development in the  
 chicken.";  
 RL Development 121:1399-1410(1995).  
 CC -1- FUNCTION: POTENT MITOGEN AND TRANSFORMING AGENT (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; 247555; CAA87635.1; -  
 DR HSP; P31371; I682.  
 DR InterPro; IPR002209; HB/F\_growthfact.  
 DR InterPro; IPR002348; IIL\_HBGF.  
 DR Pfam; PF00167; FGF; 1.  
 DR PRINTS; PR00262; IILHBGF.  
 DR ProDom; PD000831; HB/F\_growthfact; 1.  
 DR SMART; SM00442; FGF; 1.  
 DR PROSITE; PS00247; HBGF\_FGF; 1.  
 KW Growth factor; Mitogen; Signal; Glycoprotein.  
 FT SIGNAL 1 19  
 FT CHAIN 20 220  
 FT CARBOHYD 66 66 N-LINKED (GLCNAC...); (POTENTIAL).  
 FT SEQUENCE 220 AA; 25050 MW; B15D41D1551C5D5 CRC64;  
 SQ



Query Match 0.6%; Score 7; DB 1; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALL 1140  
 |||||  
 Db 6 LLLALL 12

RESULT 68  
 SDFL\_HUMAN STANDARD; PRT; 221 AA.

ID SDFL\_HUMAN  
 AC Q9HCN8; Q9BR15;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Stromal cell-derived factor 2-like protein 1 precursor (SDF2 like protein 1) (PMF1-interacting protein 8).  
 GN SDF2L1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=21092619; PubMed=1162531;  
 RA Fukuda S., Sunli M., Masuda Y., Takahashi M., Koike N., Teishima J., Yasumoto H., Itamoto T., Asahara T., Dohi K., Kamiya K.;  
 RT "Murine and human SDF2L1 is an endoplasmic reticulum stress-inducible gene and encodes a new member of the Pmc/rt protein family.";  
 RL Biochem. Biophys. Res. Commun. 280:407-414(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=uterus;  
 RA Honore B.;  
 RT "hPMP1-interacting protein 8.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=lung;  
 RA Strassberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (Potential).  
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed with high expression in testis, moderate expression in the pancreas, spleen, prostate, small intestine and colon. Very low expression is seen in brain and skeletal muscle.  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 39.  
 CC -1- SIMILARITY: CONTAINS 3 MIR DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB043007; BAB18277.1; -;  
 CC EMBL; AF277316; AAK69113.1; -;  
 CC EMBL; BC006248; AAK06248.1; -;  
 CC Genew; HGNC:10676; SDF2L1.  
 CC InterPro; IPR000866; ER\_target.  
 CC InterPro; IPR003608; MIR.  
 CC Pfam; PF02815; MIR; 3.  
 CC SMART; SM00472; MIR; 3.  
 CC PROSITE; PS00014; ER\_TARGET; 1.  
 CC Endoplasmic reticulum; Signal; Repeat.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT CHAIN 29 221 STROMAL CELL-DERIVED FACTOR 2-LIKE  
 FT PROTEIN 1.

FT DOMAIN 33 87 MIR 1.  
 FT DOMAIN 95 150 MIR 2.  
 FT DOMAIN 151 205 MIR 3.  
 FT SITE 218 221 PREVENT SECRETION FROM ER (POTENTIAL).  
 FT CONFLICT 94 94 C-> R (IN REF. 2 AND 3).  
 FT CONFLICT 162 162 L-> F (IN REF. 2 AND 3).  
 SQ SEQUENCE 221 AA; 23511 MW; 23EBE14BE9E92EP CRC64;

Query Match 0.6%; Score 7; DB 1; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1135 LLLALL 1141  
 |||||  
 Db 17 LLLALL 23

RESULT 69  
 EMBP\_MOUSE STANDARD; PRT; 223 AA.

ID EMBP\_MOUSE  
 AC Q61878;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Eosinophil granule major basic protein precursor (MBP) (Proteoglycan 2, bone marrow).  
 GN PRG2 OR MBP-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=CBA/J; TISSUE=Liver;  
 RX MEDLINE=95403975; PubMed=7673718;  
 RA Larson K.A., Horton M.A., Madden B.J., Gleich G.J., Lee N.A., Lee J.-J.;  
 RT "The identification and cloning of a murine major basic protein gene expressed in eosinophils.";  
 RL J. Immunol. 155:3002-3012(1995).  
 CC -1- FUNCTION: CYTOTOXIN AND HELMINTHOTOXIN. MBP ALSO INDUCES ANTIPARASITIC DEFENSE MECHANISMS AND IMMUNE HYPERSENSITIVITY REACTIONS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC GRANULE (CRYSTALLOID CORE) (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L46768; AAA83027.1; -;  
 CC HSSP; P13727; 1H8U.  
 CC MGD; MGI:103294; Prg2.  
 CC InterPro; IPR002352; Emaior\_bas.  
 CC InterPro; IPR001304; Lectin\_C.  
 CC Pfam; PF00059; lectin\_c; 1.  
 CC PRINTS; PR00770; EMAJORBASICP.  
 CC SMART; SM00034; CLECT; 1.  
 CC PROSITE; PS00615; C-TYPE\_LLECTIN\_1; FALSE\_NEG.  
 CC PROSITE; PS00411; C-TYPE\_LLECTIN\_2; 1.  
 CC Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;  
 KW Lectin.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT PROPEP 17 106 ACIDIC.  
 FT CHAIN 107 223 EOSINOPHIL GRANULE MAJOR BASIC PROTEIN.  
 FT DOMAIN 124 223 C-TYPE LECTIN (SHORT FORM).  
 FT DISULFD 126 221 BY SIMILARITY.

FT DISULFID 198 213 BY SIMILARITY.  
SQ SEQUENCE 223 AA; 24255 MW; 7D66D946DCAD80 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 223;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1135 LLLALLV 1141  
|||||  
Db 5 LLLALLV 11

RESULT 70  
ID BRX1 HUMAN STANDARD; PRT; 225 AA.  
AC 09HBL;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Homeobox protein BarH-like 1.  
GN BARX1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. AND VARIANT ALA-19.  
RC TISSUE=Craniofacial;  
RX MEDLINE=20453194; PubMed=10995576;  
RA Gould D.B., Walter M.A.;  
RT "Cloning, characterization, localization, and mutational screening of  
the human BARX1 gene."  
RL Genomics 68:336-342(2000).  
CC - FUNCTION: TRANSCRIPTION FACTOR, WHICH MAY BE INVOLVED IN  
CRANIOFACIAL DEVELOPMENT, IN ONTOGENESIS AND IN STOMACH  
ORGANOGENESIS. MAY HAVE A ROLE IN THE DIFFERENTIATION OF MOLARS  
FROM INCISORS. BINDS TO A REGULATORY MODULE OF THE NCAM PROMOTER.  
CC - SUBCELLULAR LOCATION: Nuclear (probable).  
CC - TISSUE SPECIFICITY: WIDELY EXPRESSED. EXPRESSED AT HIGHER LEVELS  
IN TESTIS AND HEART. DETECTED IN CRANIOFACIAL TISSUE AND ADULT  
IRIS, BUT NOT IN LYMPHOCYTES, FIBROBLASTS, CHOROIOD RETINA, RETINAL  
PIGMENT EPITHELIUM, KIDNEY, OR FETAL LIVER.  
CC - POLYMORPHISM: THE POLYMORPHISM IS NOT ASSOCIATED WITH AXENFELD-  
REIDER SYNDROME (ARS), IRIDOGONIOYOGENESIS SYNDROME (IGDS) OR  
RELATED OCULAR MALFORMATIONS.  
CC - SIMILARITY: BELONGS TO THE BAR HOMEBOX FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF213356; AAC23738.1; -  
DR HSSP; P14653; 1872.  
DR Genew; HGNC:955; BARX1.  
DR MIM; 603260; -  
DR InterPro; IPR000047; HTH\_repressor.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PRO0024; HOMEBOX.  
DR PRINTS; PRO0031; HTHREPRESSR.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KM Homeobox; Transcription; DNA-binding; Nuclear protein; Polymorphism.  
FT DNA\_BIND 113 172 HOMEBOX.  
FT DOMAIN 15 24 POLY-ALA.  
FT VARIANT 19 19 T->A.  
/FTID=VAR\_010927.

SQ SEQUENCE 225 AA; 24061 MW; A7907BB466F3393 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 225;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1019 LTPPGP 1025  
|||||  
Db 6 LTPPGP 12

RESULT 71  
ID EMBP RAT STANDARD; PRT; 227 AA.  
AC 063189;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Eosinophil granule major basic protein precursor (MBP).  
GN PRG2 OR MBP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Bone marrow;  
RX MEDLINE=96138543; PubMed=8547309; Misawa S., Isobe Y., Hayashi H.,  
RA Nittoh T., Watanabe M., Okayama H.,  
RA Mue S., Ohuchi K.;  
RT "Cloning of cDNA for rat eosinophil major basic protein."  
RL Biochim. Biophys. Acta 1264:261-264(1995).  
CC - FUNCTION: CYTOTOXIN AND HELMINTHOTOXIN. MBP ALSO INDUCES  
NONCYTOLYTIC HISTAMINE RELEASE FROM BASOPHILS. IT IS INVOLVED IN  
ANTIPARASITIC DEFENSE MECHANISMS AND IMMUNE HYPERSENSITIVITY  
REACTIONS (BY SIMILARITY).  
CC - SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC  
GRANULE (CRYSTALLOID CORE) (BY SIMILARITY).  
CC - SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D50568; BAA09129.1; -  
DR HSSP; P13727; 1H8U.  
DR InterPro; IPR002352; Emaior\_baslc.  
DR InterPro; IPR001304; Lectin\_C.  
DR Pfam; PF00059; lectin\_c; 1.  
DR PRINTS; PRO0770; EMAJORBASICP.  
DR SMART; SM00034; CLECT; 1.  
DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
DR PROSITE; PS00041; C-TYPE LECTIN\_2; 1.  
KM Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;  
Lectin.  
FT SIGNAL 1 16 POTENTIAL.  
FT PROPEP 17 110 ACIDIC (BY SIMILARITY).  
FT CHAIN 111 227 EOSINOPHIL GRANULE MAJOR BASIC PROTEIN.  
FT DOMAIN 128 227 C-TYPE LECTIN (SHORT FORM).  
FT DISULFID 130 225 BY SIMILARITY.  
FT DISULFID 202 217 BY SIMILARITY.  
FT DOMAIN 59 68 POLY-GLU.  
SQ SEQUENCE 227 AA; 25129 MW; 34F2B514090DE82 CRC64;  
Query Match 0.6%; Score 7; DB 1; Length 227;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1135 LLLALLV 1141

```

Db          5 |||||
           5 LELALLV 11

RESULT 72
NH40 CAEEL STANDARD; PRT; 231 AA.
ID NH40 CAEEL
AC Q22127;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclear hormone receptor family member nhr-40.
GN NHR-40 OR 10366.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Talc A.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ORPHAN NUCLEAR RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, U40940; AAA81704.2; -.
DR HSSP; P20393; 1A6Y.
DR WormPep; T03G6.2; CE28074.
DR InterPro; IPR000356; Hormone_rec_lig.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1_1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW zinc-finger.
FT DNA_BIND 15 77 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 15 35 C4-TYPE.
FT ZN_FING 51 72 C4-TYPE.
SQ SEQUENCE 231 AA; 26781 MW; EF68F26394E94F2 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1157 EEEKREE 1163
DB 103 EEEKREE 109

RESULT 73
Y867 RICPR STANDARD; PRT; 231 AA.
ID Y867 RICPR
AC Q9ZC39;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Hypothetical protein RP867.
GN RP867.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiinae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E.; Zomorodipour A.; Andersson J.O.;
RA Sickeritz-Ponten T.; Alsmark U.C.M.; Podowski R.M.; Naeslund A.K.;
RA Eriksson A.-S.; Winkler H.H.; Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG, TO P.DENITRIFICANS HYPOTHETICAL 25.1 KDa
CC PROTEIN IN COBV 5'REGION (AC P29938).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, AJ235273; CAA15291.1; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 66 86 POTENTIAL.
SQ SEQUENCE 231 AA; 26472 MW; 4B3FE686FA02343 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 LQMSILE 114
DB 209 LQMSILE 215

RESULT 74
EMBL_CAVPO STANDARD; PRT; 233 AA.
ID EMBL_CAVPO
AC P22032;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Eosinophil granule major basic protein 1 precursor (MBP-1).
KW MBP.
GN Cavia porcellus (Guinea pig).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 115-162.
RC TISSUE=Eosinophil; PubMed=1705901;
RX MEDLINE=91160746; PubMed=1705901;
RA Aoki I.; Shindoh Y.; Nishida T.; Nakai S.; Hong Y.-M.; Mio M.;
RA Saito T.; Taseka K.;
RT "Sequencing and cloning of the cDNA of guinea pig eosinophil major
RT basic protein."
RL FEBS Lett. 279:330-334(1991).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=94092714; PubMed=8268206;
RA Hashimoto Y.; Nagaoaka I.; Yamashita T.;
RT "Purification of the antibacterial fragments of guinea-pig major
RT basic protein."
RL Biochim. Biophys. Acta 1203:236-242(1993).

```

```

CC -!- FUNCTION: MBP MAY PLAY SOME IMPORTANT ROLES IN THE ALLERGIC
CC REACTIONS AND INFLAMMATIONS. SINCE MBP IS CAPABLE OF RELEASING
CC HISTAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF
CC BRONCHIAL TUBES. ANTIPARASITIC AND ANTI-BIOTIC.
CC -!- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC
CC GRANULE (CRYSTALLOID CORE).
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90251; BAA14291.1; -.
DR PIR; S13625; S13625.
DR HSSP; P13727; 1H8U.
DR InterPro; IPR002352; Emaior basic.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; Lectin_C; 1.
DR PRINTS; PR00770; EMAIORBASIC.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;
KW Lectin; Multigene family.
FT SIGNAL 1 15 POTENTIAL.
FT PROPEP 16 114 ACIDIC.
FT CHAIN 115 233 EOSINOPHIL GRANULE MAJOR BASIC PROTEIN 1.
FT DOMAIN 132 233 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 134 231 BY SIMILARITY.
FT DISULFID 208 223 BY SIMILARITY.
SQ SEQUENCE 233 AA; 26268 MW; C8D5E96D927C56C8 CRC64;

```

```

Query Match 0.6%; Score 7; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1134 LLLALL 1140
DB 4 LLLALL 10

```

```

RESULT 75
COMB_CLOAB
ID COMB_CLOAB STANDARD; PRT; 235 AA.
AC Q97E82;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable 2-phosphosulfolactate phosphatase (EC 3.1.3.-).
GN COMB OR CAC3233.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=146286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabatne F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- CATALYTIC ACTIVITY: 2-phosphosulfolactate = sulfolactate +
CC phosphate.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SIMILARITY: BELONGS TO THE COMB FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007819; AAK81168.1; -.
KW Hydrolase; Magnesium; Complete proteome.
SQ SEQUENCE 235 AA; 26144 MW; ABBFB7DF07161EBD CRC64;

```

```

Query Match 0.6%; Score 7; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1125 ILIGSVL 1131
DB 118 ILIGSVL 124

```

```

Search completed: July 16, 2003, 08:32:39
Job time : 19 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 08:30:30 ; Search time 29 Seconds  
(without alignments)  
3868.582 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167  
Sequence: 1 MELPVTHTLPLVFLTGLC.....GFFAHKKIPEERREKLEQ 1167

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	1.1	272	2 A55348	integrin alpha-1 -
2	13	1.1	1151	2 A45226	integrin alpha-1 c
3	13	1.1	1180	2 A35854	integrin alpha-1 c
4	11	0.9	191	2 I47230	VLA-2 protein - pi
5	11	0.9	687	2 C71849	carbon starvation
6	11	0.9	687	2 H64665	carbon starvation
7	11	0.9	1170	2 I45914	integrin alpha 2 s
8	11	0.9	1178	2 S44142	VLA-2 protein homo
9	11	0.9	1181	2 A33998	integrin alpha-2 c
10	10	0.9	1179	2 A53213	integrin alpha-E c
11	9	0.8	84	2 B95310	hypothetical prote
12	9	0.8	104	2 PH0141	hypothetical prote
13	9	0.8	236	2 T12766	hypothetical prote
14	9	0.8	1170	2 I56126	probable lipoprote
15	9	0.8	1170	2 S03308	lymphocyte functio
16	8	0.7	42	2 A60822	cell surface glyco
17	8	0.7	20	2 A34259	cytochrome P450 PB
18	8	0.7	76	2 A43537	cytochrome P450mt4
19	8	0.7	76	2 I53107	heat-stable antige
20	8	0.7	80	2 A48966	CD24 precursor - r
21	8	0.7	103	2 AE0856	B cell surface ant
22	8	0.7	127	2 B75301	conserved hypothet
23	8	0.7	136	2 G82058	hypothetical prote
24	8	0.7	141	2 S15785	ribosomal protein
25	8	0.7	169	2 AB1010	heat-stable antige
26	8	0.7	185	2 H96977	formylmethionine d
27	8	0.7	234	1 S15102	probable membrane
28	8	0.7	240	2 E72629	eosinophil major b
29	8	0.7	251	2 A81345	hypothetical prote
					histidine-binding
30	8	0.7	254	2 B55904	C 3.4.25.1 proteas
31	8	0.7	259	2 T37985	proteasome compone
32	8	0.7	264	2 S22090	catechol O-methylt
33	8	0.7	281	2 S18541	hypothetical prote
34	8	0.7	298	2 F71556	probable had hydro
35	8	0.7	309	2 C65217	aliose kinase (EC
36	8	0.7	360	2 A85016	hypothetical prote
37	8	0.7	367	2 S19172	cytochrome P450 2B
38	8	0.7	383	2 T31771	hypothetical prote
39	8	0.7	415	2 G83568	probable permease
40	8	0.7	433	2 AC2275	hypothetical prote
41	8	0.7	444	2 T01721	hypothetical prote
42	8	0.7	491	1 OARTPB	cytochrome P450 2B
43	8	0.7	491	1 OARTP2	cytochrome P450 2B
44	8	0.7	500	2 B31047	testosterone 16alp
45	8	0.7	572	2 I39311	Kruppel-type zinc
46	8	0.7	604	2 T36917	glycoprotein Iib -
47	8	0.7	607	2 S60658	legumain - Gnaetun g
48	8	0.7	614	2 A69845	Na+/H+ antiporter
49	8	0.7	685	2 AC0527	ferrichrome transp
50	8	0.7	767	2 G96661	hypothetical prote
51	8	0.7	849	2 E83349	probable ClpA/B-ty
52	8	0.7	942	1 J01674	protein kinase TMK
53	8	0.7	959	1 B60017	outer capsid prote
54	8	0.7	969	2 T15446	hypothetical prote
55	8	0.7	1037	2 A60163	glycoprotein Iib -
56	8	0.7	1039	2 A34269	integrin alpha-2b
57	8	0.7	1146	2 S40311	integrin - fruit f
58	8	0.7	1153	2 RMHUIB	cell surface glyco
59	8	0.7	1163	1 RMHUIC	cell surface glyco
60	8	0.7	1500	2 G84922	hypothetical prote
61	8	0.7	2109	2 I38414	transcription fact
62	8	0.7	2222	2 T13924	sdh protein - fru
63	8	0.7	3176	1 CGHUI3	collagen alpha 3(V
64	7	0.6	47	2 D81666	hypothetical prote
65	7	0.6	73	2 C65069	hypothetical prote
66	7	0.6	74	2 T10456	demaesepine B3 pr
67	7	0.6	76	2 I51527	integrin alpha 5 s
68	7	0.6	77	2 I51529	integrin alpha V-1
69	7	0.6	84	2 D70640	hypothetical prote
70	7	0.6	85	2 G82776	ACP xP0672 [import
71	7	0.6	86	2 B10993	hypothetical prote
72	7	0.6	86	2 G85938	hypothetical prote
73	7	0.6	89	2 I49515	B144 protein B - m
74	7	0.6	94	2 T03285	yeast-specific pr
75	7	0.6	97	1 A64146	hypothetical prote
76	7	0.6	100	2 F72587	hypothetical prote
77	7	0.6	107	1 IFCH	insulin precursor
78	7	0.6	108	2 S40149	integrin alpha-7C
79	7	0.6	109	2 H72713	hypothetical prote
80	7	0.6	110	2 S30268	protein hdea precu
81	7	0.6	110	2 F91177	protein hdea precu
82	7	0.6	110	2 G66023	protein hdea precu
83	7	0.6	111	2 T36386	probable Lsr2-Like
84	7	0.6	115	2 D75434	T-cell receptor be
85	7	0.6	119	2 B32578	hypothetical prote
86	7	0.6	127	2 S76755	hypothetical prote
87	7	0.6	129	2 F72488	hypothetical prote
88	7	0.6	130	2 D83305	hypothetical prote
89	7	0.6	132	2 B81215	conserved hypothet
90	7	0.6	133	2 H75386	hypothetical prote
91	7	0.6	137	2 P00502	hypothetical prote
92	7	0.6	139	2 A38612	insulin-like growt
93	7	0.6	142	1 S59391	hypothetical prote
94	7	0.6	142	2 UT0573	retinoic acid-indu
95	7	0.6	142	2 G87647	hypothetical prote
96	7	0.6	143	2 T43180	hypothetical prote
97	7	0.6	143	2 T40426	hypothetical prote
98	7	0.6	144	2 A44585	globin - sea hare
99	7	0.6	145	2 B72724	hypothetical prote
100	7	0.6	148	2 D75090	hypothetical prote

## ALIGNMENTS

## RESULT 1

A55348  
Integrin alpha-1 - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 10-Feb-1995 #sequence\_revision 10-Feb-1995 #text\_change 17-Mar-1999  
C:Accession: A55348  
R:Kern, A.; Brissewitz, R.; Bank, I.; Marcantonio, E.E.  
J. Biol. Chem. 269, 22811-22816, 1994  
A:Title: The role of the I domain in ligand binding of the human integrin alpha-1b beta-1.  
A:Reference number: A55348; PMID:94357930; PMID:7521332  
A:Accession: A55348  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-272 <KER>  
A:Cross-references: GB:U10114  
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology  
F:55-230/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match  
Best Local Similarity 100.0%; Score 13; DB 2; Length 272;  
Pred. No. 0.00016;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181  
DB 59 VIVLDGNSIYPM 71

## RESULT 2

A45226  
Integrin alpha-1 chain - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2000  
C:Accession: A45226  
R:Brisewitz, R.; Epstein, M.R.; Marcantonio, E.E.  
J. Biol. Chem. 268, 2989-2996, 1993  
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.  
A:Reference number: A45226; PMID:93155124; PMID:8428973  
A:Accession: A45226  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1151 <BRI>  
A:Experimental source: hepatoblastoma cell line HepG2  
A:Note: sequence extracted from NCBI backbone (NCBIP:124326)  
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology  
F:142-317/Domain: von Willebrand factor type A repeat homology <VMA1>

Query Match  
Best Local Similarity 100.0%; Score 13; DB 2; Length 1151;  
Pred. No. 0.00052;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181  
DB 146 VIVLDGNSIYPM 158

## RESULT 3

A35854  
Integrin alpha-1 chain precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 23-Oct-1990 #sequence\_revision 13-Sep-1991 #text\_change 20-Sep-1999  
C:Accession: A35854; S11243  
R:Ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto, J. Cell Biol. 111, 709-720, 1990  
A:Title: Molecular cloning of the rat integrin alpha-1 subunit: a receptor for laminin A  
A:Reference number: A35854; PMID:90338125; PMID:2380249  
A:Accession: A35854  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1180 <IGN>  
A:Cross-references: GB:X52140; NID:956493; PIDN:CAA36384.1; PID:956494

C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology  
C:Keywords: cell adhesion; cytoskeleton; transmembrane protein  
F:170-345/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match  
Best Local Similarity 100.0%; Score 13; DB 2; Length 1180;  
Pred. No. 0.00053;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181  
DB 174 VIVLDGNSIYPM 186

## RESULT 4

I47230  
VLA-2 protein - pig (fragment)  
A:Alternate names: glycoprotein Ia/IIa  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 20-Sep-1999  
C:Accession: I47230; S21518  
R:Bahou, W.F.; Potter, C.L.; Mirza, H.  
Blood 84, 3734-3741, 1994  
A:Title: The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific recognition  
A:Reference number: I47230; PMID:95036279; PMID:7949129  
A:Accession: I47230  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-191 <BAH>  
A:Cross-references: EMBL:Z12137; NID:92158; PIDN:CAA78125.1; PID:92159  
C:Superfamily: unassigned collagens  
C:Keywords: glycoprotein

Query Match  
Best Local Similarity 100.0%; Score 11; DB 2; Length 191;  
Pred. No. 0.013;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282  
DB 92 VVVTGESHHDG 102

## RESULT 5

C71849  
Carbon starvation protein A homolog jhp1095 [similarity] - Helicobacter pylori (strain J  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 18-Aug-2000  
C:Accession: C71849  
R:Alm, R.A.; Lang, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Mille, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; PMID:99120557; PMID:9923682  
A:Accession: C71849  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-687 <ARN>  
A:Cross-references: GB:AE001537; GB:AE001439; NID:94155687; PIDN:AAD06677.1; PID:9415569  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp1095  
C:Superfamily: carbon starvation protein

Query Match  
Best Local Similarity 0.9%; Score 11; DB 2; Length 687;  
Pred. No. 0.037;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 LMIILGSVLGG 1133  
DB 118 LMIILGSVLGG 128

## RESULT 6

H64665  
Carbon starvation protein A homolog HP1168 [similarity] - Helicobacter pylori (strain 26  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 18-Aug-2000  
C:Accession: H64665  
R:Tombl, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodok, A.; McKenne  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, L.  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Kark, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: H64665  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-687 <TOM>  
A:Cross-references: GB:AE000622; GB:AE000511; NID:g2314317; PIDN:AAD08212.1; PID:g231432  
C:Superfamily: Carbon starvation protein

Query Match 0.9%; Score 11; DB 2; Length 687;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 LMIIGSVLGG 1133  
|||||  
Db 118 LMIIGSVLGG 128

## RESULT 7

I45914  
Integrin alpha 2 subunit - bovine (fragment)  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 20-Sep-1999  
C:Accession: I45914  
R:Kamata, T.; Puzon, W.; Takada, Y.  
J. Biol. Chem. 269, 9659-9663, 1994  
A:Title: Identification of putative ligand binding sites within the I-domain of integrin  
A:Reference number: A54402; MUID:94193647; PMID:7511592  
A:Accession: I45914  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1170 <KAM>  
A:Cross-references: GB:I25886; NID:9439695; PIDN:AAB59235.1; PID:9439696  
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology  
F:161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 0.9%; Score 11; DB 2; Length 1170;  
Best Local Similarity 100.0%; Pred. No. 0.058;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282  
|||||  
Db 268 VVVTGESHHDG 278

## RESULT 8

S44142  
VLA-2 protein homolog - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Sep-1999  
C:Accession: S44142  
R:Edelman, J.M.; Chan, B.M.; Uthyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latze  
submitted to the EMBL Data Library, January 1994  
A:Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not  
A:Reference number: S44142  
A:Accession: S44142  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1178 <EDE>  
A:Cross-references: EMBL:229987; NID:9473098; PIDN:CAA82877.1; PID:9473099  
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology  
F:163-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 0.9%; Score 11; DB 2; Length 1178;  
Best Local Similarity 100.0%; Pred. No. 0.058;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282  
|||||  
Db 276 VVVTGESHHDG 286

## RESULT 9

A33998  
Integrin alpha-2 chain precursor - human  
N:Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 ch.  
C:Species: Homo sapiens (man)  
C:Date: 30-Mar-1990 #sequence\_revision 18-Sep-1992 #text\_change 21-Jul-2000  
C:Accession: A33998; B56793; A53117  
R:Takada, Y.; Hemler, M.E.  
J. Cell. Biol. 109, 397-407, 1990  
A:Title: The primary structure of the VLA-2/collagen receptor alpha(2)-subunit (platelet  
A:Reference number: A33998; MUID:89308879; PMID:2545729  
A:Accession: A33998  
A:Molecule type: mRNA  
A:Residues: 1-1181 <TK>  
A:Cross-references: GB:X17033; NID:933906; PIDN:CAA34894.1; PID:933907  
A:Note: the authors translated the codon GAT for residue 802 as Gln, GTC for residue 803  
R:Camel, B.; Parentier, S.; Leung, L.L.; McGregor, J.L.  
Biochem. J. 279, 419-425, 1991  
A:Title: Separation of important new platelet glycoproteins (GPIa, GPIC, GPIC\*, GP1a an  
A:Reference number: A56793; MUID:92061944; PMID:1955640  
A:Accession: B56793  
A:Molecule type: protein  
A:Residues: 30-43 <CAT>  
A:Experimental source: platelet  
R:Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Tsung, Y.L.; Gafford, A.  
J. Biol. Chem. 269, 463-469, 1994  
A:Title: The human alpha-2 integrin gene promoter. Identification of positive and negati  
A:Reference number: A53117; MUID:94103255; PMID:8276836  
A:Accession: A53117  
A:Molecule type: DNA  
A:Residues: 1-16, 'V', 18-21 <ZUT>  
A:Cross-references: GB:I24121; NID:9400342; PIDN:AAA16619.2; PID:94583535  
A:Note: authors translated the codon GTA for residue 17 as Leu  
C:Genetics:  
A:Gene: GDB:ITGA2; CD49B  
A:Cross-references: GDB:128031; OMIM:192974  
A:Map position: 5q11.1-5q11.2

C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology  
C:Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-1133/Domain: extracellular #status predicted <EXT>  
F:172-347/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1134-1154/Domain: transmembrane #status predicted <TM>  
F:1155-1181/Domain: intracellular #status predicted <CT>  
F:105,112,343,432,460,475,699,1057,1074,1081/Binding site: carbohydrate (Asn) (covalent)

Query Match 0.9%; Score 11; DB 2; Length 1181;  
Best Local Similarity 100.0%; Pred. No. 0.058;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282  
|||||  
Db 279 VVVTGESHHDG 289

## RESULT 10

A53213  
Integrin alpha-E chain - human  
C:Species: Homo sapiens (man)  
C:Date: 19-Oct-1995 #sequence\_revision 31-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A53213  
R:Shaw, S.K.; Ceppek, K.L.; Murphy, E.A.; Russell, G.J.; Brenner, M.B.; Parker, C.M.  
J. Biol. Chem. 269, 6016-6023, 1994  
A:Title: Molecular cloning of the human mucosal lymphocyte integrin alpha(E) subunit. Un

A:Reference number: A53213; MUID:94164962; PMID:8119947  
 A:Accession: A53213  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1179 <SHA>  
 A:Cross-references: GB:125851; NID:9457244; PID:9457245  
 C:Genetics:  
 A:Gene: GDB:ITGAE  
 A:Cross-references: GDB:330801  
 A:Map position: 17p13  
 C:Superfamily: cell surface glycoprotein CD11b, von Willebrand factor type A repeat homology <WMA3>  
 F:199-371/Domain: von Willebrand factor type A repeat homology <WMA3>

Query Match 0.9%; Score 10; DB 2; Length 1179;  
 Best Local Similarity 100.0%; Pred. No. 0.61;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYGSELCP 492  
 |||||  
 Db 511 GSYGSELCP 520

RESULT 11  
 B95310  
 hypotetical protein SMA0726 [imported] - Sinorhizobium meliloti (strain 1021) magaplasma  
 C:Species: Sinorhizobium meliloti  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: B95310  
 R:Karnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bower, J.; Katman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: B95310  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-84 <KUN>  
 A:Cross-references: GB:AE006469; PIDN:AAK65044.1; PID:914523476; GSPDB:GN00165  
 A:Experimental source: strain 1021, megaplasmid pSymA  
 R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chai, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Leilaure, H.; P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: SMA0726  
 A:Genome: plasmid

Query Match 0.8%; Score 9; DB 2; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 0.72;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 RLSPRLRL 729  
 |||||  
 Db 58 RLSPRLRL 66

RESULT 12  
 PH0141  
 integrin alpha chain - chicken (fragments)  
 C:Species: Gallus gallus (chicken)  
 C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 16-Apr-1999  
 C:Accession: PH0141  
 R:Syfrid, J.; Mann, K.; Paulsson, M.  
 Exp. Cell Res. 194, 165-173, 1991  
 A:Title: An abundant chick gizzard integrin is the avian alpha1-beta1 integrin heterodimer  
 A:Reference number: PH0141; MUID:91224161; PMID:1851093  
 A:Accession: PH0141  
 A:Molecule type: protein

A:Residues: 1-104 <SYF>  
 A:Experimental source: gizzard  
 C:Comment: Integrins are a family of cellular adhesion receptors that mediate interaction  
 C:Superfamily: unassigned collagens  
 C:Keywords: cell adhesion

Query Match 0.8%; Score 9; DB 2; Length 104;  
 Best Local Similarity 100.0%; Pred. No. 0.86;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 GEQIGSYFG 487  
 |||||  
 Db 43 GEQIGSYFG 51

RESULT 13  
 T12766  
 probable lipoprotein yoxB - Bacillus subtilis phage SPhC2  
 C:Species: Bacillus subtilis phage SPhC2  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 15-Oct-1999  
 C:Accession: T12766; F69907  
 R:Razarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.  
 submitted to the EMBL Data Library, August 1997  
 A:Description: The complete nucleotide sequence of the Bacillus subtilis SPhC2 prophage  
 A:Reference number: 217583  
 A:Accession: T12766  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-236 <LAZ>  
 A:Cross-references: EMBL:AF020713; NID:93025478; PID:93025480; PIDN:AAI12975.1  
 R:Kunat, F.; Ogatawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter, C.; Bron, S.; Bruni, S.; Brusch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Cho, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Hentut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Y.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue, V. M.; Ogata, K.; Ogatawara, A.; Oudaga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle, Reger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadie, Y.; Sato, T.; Scanlon, A.; Authors: Scheidt, S.; Schreier, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seror, A.; Whalers, P.; Wypat, A.; Tanaka, T.; Terpetra, F.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Whalers, P.; Wypat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumelein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: F69907  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-236 <KUN>  
 A:Cross-references: GB:259115; GB:AL009126; NID:92634478; PIDN:CAB14083.1; PID:el183612;  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yoxB

Query Match 0.8%; Score 9; DB 2; Length 236;  
 Best Local Similarity 100.0%; Pred. No. 1.7;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 PEEKREK 1164  
 |||||  
 Db 209 PEEKREK 217

RESULT 14  
 I56126  
 lymphocyte function-associated molecule-1-alpha - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 20-Aug-1999  
 C:Accession: I56126  
 R:Kautmann, Y.; Tseng, E.; Springer, T.A.  
 J. Immunol. 147, 369-374, 1991  
 A:Title: Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit a



A:Reference number: 156126; MUID:91268576; PMID:2051027  
A:Accession: 156126  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1163 <RES>  
A:Cross-references: GB:M60778; NID:9198785; PIDN:AAA9426.1; PID:9198786  
C:Gene: LFA-1  
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology F.151-315/Domain: von Willebrand factor type A repeat homology <VWA1>  
Query Match 0.8%; Score 9; DB 2; Length 1163;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 566 DVAVGAPLE 574  
DB 536 DVAVGAPLE 544  
RESULT 15  
S03308  
cell surface glycoprotein CD11a precursor - human  
M:Alternate names: leukocyte adhesion glycoprotein LFA-1 alpha chain; leukocyte function C:Species: Homo sapiens (man)  
C:Date: 28-Feb-1990 #sequence revision 28-Feb-1990 #text\_change 20-Aug-1999  
C:Accession: S03308; A47458; A47565; S36044  
R:Larsen, R.S.; Corbi, A.L.; Berman, L.; Springer, T.  
J:Cell Biol. 108, 703-712, 1989  
A:Title: Primary structure of the leukocyte function-associated molecule-1 alpha subunit  
A:Reference number: S03308; MUID:89139587; PMID:2537322  
A:Accession: S03308  
A:Molecule type: mRNA  
A:Residues: 1-1170 <LAR>  
A:Cross-references: EMBL:Y00796; NID:931421; PIDN:CAA68747.1; PID:931422  
A>Note: sequence extracted from NCBI backbone (NCBIN:130862; NCBI:130863)  
R:Shelley, C.S.; Partridge, O.C.; Arnaout, M.A.  
Proc. Natl. Acad. Sci. U.S.A. 90, 5364-5368, 1993  
A:Title: Identification of cell-specific and developmentally regulated nuclear factors  
A:Reference number: A47565; MUID:93281759; PMID:8099450  
A:Accession: A47565  
A:Molecule type: DNA  
A:Residues: 1-20 <SHE>  
R:Nueda, A.; Lopez-Cabrera, M.; Vaz, A.; Corbi, A.L.  
J. Biol. Chem. 268, 19305-19311, 1993  
A:Title: Characterization of the CD11a (alphaL, LFA-1alpha) integrin gene promoter.  
A:Reference number: A48759; MUID:93374910; PMID:8103515  
A:Accession: A48759  
A:Molecule type: DNA  
A:Residues: 1-20 <NUE>  
A:Cross-references: EMBL:Z22804; NID:9311405; PIDN:CAA80461.1; PID:9311406  
C:Gene: GDB: ITGAL; CD11A  
A:Cross-references: GDB:119757; OMIM:153370  
A:Map position: 16p11.2-16p11.2  
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology F.1-25/Domain: signal sequence status predicted <SIG>  
F.1-25/Domain: signal sequence status predicted <SIG>  
F.26-1170/Product: leukocyte adhesion glycoprotein LFA-1 alpha chain #status predicted <VWA2>  
F.154-317/Domain: von Willebrand factor type A repeat homology <VWA2>  
Query Match 0.8%; Score 9; DB 2; Length 1170;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 566 DVAVGAPLE 574  
DB 538 DVAVGAPLE 546  
RESULT 16  
A60822  
cytochrome P450 PB-3a - rat (fragment)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text\_change 11-Jun-1999  
C:Accession: A60822; 155191  
R:Ameliaz, Z.; Nardonne, J.F.; Wolf, C.R.; Robertson, L.W.; Oesch, F.  
Biochem. Pharmacol. 37, 3245-3249, 1988  
A:Title: Effect of nutritional imbalances on cytochrome P-450 isozymes in rat liver.  
A:Reference number: A60822; MUID:88293549; PMID:3041969  
A:Accession: A60822  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-20 <AMB>  
R:Hashimoto, T.; Matsumoto, T.; Nishizawa, M.; Kawabata, S.  
J. Biochem. 103, 487-492, 1988  
A:Title: A mutant rat strain deficient in induction of a phenobarbital-inducible form of C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
A:Reference number: 155191; MUID:88273074; PMID:2839467  
A:Accession: 155191  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-20 <RES>  
A:Cross-references: GB:D00250; NID:9220727; PIDN:BA00181.1; PID:9220728  
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
C:Keywords: electron transfer; heme; monooxygenase; oxidoreductase; transmembrane protein  
Query Match 0.7%; Score 8; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1134 LLLALLLV 1141  
DB 6 LLLALLLV 13  
RESULT 17  
A34259  
cytochrome P450mt4, phenobarbital-inducible, mitochondrial, hepatic - rat (fragment)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text\_change 05-Mar-1999  
C:Accession: A34259  
R:Shayiq, R.M.; Avadhani, N.G.  
Biochemistry 29, 866-873, 1990  
A:Title: A phenobarbital-inducible hepatic mitochondrial cytochrome P-450 immunochromatography  
A:Reference number: A34259; MUID:90254127; PMID:2340279  
A:Accession: A34259  
A:Molecule type: protein  
A:Residues: 1-42 <SHA>  
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
C:Keywords: electron transfer; heme; liver; mitochondrion; monooxygenase; oxidoreductase  
Query Match 0.7%; Score 8; DB 2; Length 42;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1134 LLLALLLV 1141  
DB 6 LLLALLLV 13  
RESULT 18  
A43537  
heat-stable antigen M1/69-J11d precursor - mouse  
M:Alternate names: CD24 protein; necladixin  
C:Species: Mus musculus (house mouse)

C>Date: 06-Nov-1992 #sequence, revision 06-Nov-1992 #text, change 31-Jan-2000  
 C/Accession: A43537; 148287; S15784; S15783; S43709; S32240; S33129  
 R/Kay, R.; Takei, F.; Humphries, R.K.  
 J. Immunol. 145, 1952-1959, 1990  
 A>Title: Expression cloning of a cDNA encoding M1/69-J11d heat-stable antigens.  
 A/Reference number: A43537; MUID:90361906; PMID:2118158  
 A/Accession: A43537  
 A/Molecule type: mRNA  
 A/Residues: 1-76 <KX>  
 A/Cross-references: GB:M58661; NID:g198985; PIDN:AA39481.1; PID:g198986  
 R/Menger, R.H.; Rochelle, J.M.; Seidin, M.F.; Kohler, G.; Nielsen, P.J.  
 J. Biol. Chem. 268, 23345-23352, 1993  
 A>Title: The heat stable antigen (mouse CD24) gene is differentially regulated but has a  
 A/Reference number: A48876; MUID:94043137; PMID:8226859  
 A/Accession: 148287  
 A/Status: translation not shown; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-76 <RES>  
 A/Cross-references: EMBL:X72910; NID:g296466; PIDN:CAA51415.1; PID:g296467  
 R/Menger, R.H.; Ayane, M.; Bose, R.; Koehler, G.; Nielsen, P.J.  
 Eur. J. Immunol. 21, 1039-1046, 1991  
 A>Title: The genes for a mouse hematopoietic differentiation marker called the heat-stab  
 A/Reference number: S15783; MUID:91209380; PMID:2019286  
 A/Accession: S15784  
 A/Molecule type: DNA  
 A/Residues: 1-76 <WE3>  
 A/Cross-references: EMBL:X56469; NID:g51439; PIDN:CAA39841.1; PID:g51440  
 A/Accession: S15783  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 32-76 <WE2>  
 A/Cross-references: EMBL:X53825  
 R/Nielsen, P.J.  
 submitted to the EMBL Data Library, July 1990  
 A/Reference number: S19111  
 A/Accession: S43709  
 A/Molecule type: mRNA  
 A/Residues: 1-76 <NIE>  
 A/Cross-references: EMBL:X53825; NID:g55441; PIDN:CAA37822.1; PID:g55442  
 C/Genetics:  
 A/Gene: CD24  
 A/Map position: 10  
 A/Intons: 23/3  
 C/Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati  
 F:1-26/Domain: signal sequence #status predicted <Sig>  
 F:27-56/Product: heat-stable antigen M1/69-J11d #status predicted <MAT>  
 F:57-76/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
 F:56/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form)

Query Match 0.7%; Score 8; DB 2; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GULLALL 1140  
 DB 12 GULLALL 19

RESULT 19  
 153107  
 CD24 precursor - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 02-Aug-1996 #sequence, revision 02-Aug-1996 #text, change 05-Nov-1999  
 C/Accession: I53107; S25146  
 R/Shirasawa, T.; Akashi, T.; Sakamoto, K.; Takahashi, H.; Maruyama, N.; Hirokawa, K.  
 Dev. Dyn. 198, 1-13, 1993  
 A>Title: Gene expression of CD24 core peptide molecule in developing brain and developin  
 A/Reference number: I53107; MUID:94122434; PMID:8292828  
 A/Accession: I53107  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-76 <RES>  
 A/Cross-references: EMBL:211663; NID:g55901; PIDN:CAA7731.1; PID:g55902

C/Keywords: phosphatidylinositol linkage

Query Match 0.7%; Score 8; DB 2; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GULLALL 1140  
 DB 12 GULLALL 19

RESULT 20  
 A48996  
 B cell surface antigen CD24 precursor - human  
 N/Alternate names: cluster-w4 antigen; signal transducer CD24  
 C/Species: Homo sapiens (man)  
 C/Date: 19-Dec-1993 #sequence, revision 02-Aug-1996 #text, change 21-Jul-2000  
 C/Accession: I56114; A48996; I54201; I72676  
 R/Kay, R.; Rosten, P.M.; Humphries, R.K.  
 J. Immunol. 147, 1412-1416, 1991  
 A>Title: CD24, a signal transducer modulating B cell activation responses, is a very sho  
 A/Reference number: I56114; MUID:91332458; PMID:1831224  
 A/Accession: I56114  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-80 <RES>  
 A/Cross-references: GB:L33930; NID:g500848; PID:g500849; GB:M58664; NID:g180167; PID:g18  
 R/Jackson, D.; Walpel, R.; Weber, E.; Bell, J.; Stahel, R.A.  
 Cancer Res. 52, 5264-5270, 1992  
 A>Title: CD24, a signal-transducing molecule expressed on human B cells, is a major surf  
 A/Reference number: A48996; MUID:93007871; PMID:1327504  
 A/Accession: A48996  
 A/Molecule type: mRNA  
 A/Residues: 1-56, 'V', 58-80 <JAC>  
 A/Cross-references: GB:X63397; GB:S44888; NID:g396167; PIDN:CAA49195.1; PID:g396168  
 A/Experimental source: small cell lung carcinoma line SW2  
 A/Note: sequence extracted from NCBI Backbone (NCBIN:114635, NCBI:P:114635)  
 R/Hough, M.R.; Rosten, P.M.; Sexton, T.L.; Kay, R.; Humphries, R.K.  
 Genomics 22, 154-161, 1994  
 A>Title: Mapping of CD24 and homologous sequences to multiple chromosomal loci.  
 A/Reference number: I54201; MUID:95048364; PMID:7959762  
 A/Accession: I54201  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 5-11, 'W', 13-43, 'T', 45-80 <RE2>  
 A/Cross-references: GB:S75311; NID:g833885; PIDN:AA014170.1; PID:g4261870  
 C/Genetics:  
 A/Gene: GDB:CD24  
 A/Cross-references: GDB:383795; OMIM:600074  
 A/Map position: 6q21-6q21  
 C/Keywords: B-cell; blocked carboxyl end; glycoprotein; lipoprotein; phosphatidylinosito  
 F:1-22/Domain: signal sequence #status predicted <Sig>  
 F:22-67/Product: B cell surface antigen CD24 #status predicted <MAT>  
 F:68-80/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
 F:67/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form)

Query Match 0.7%; Score 8; DB 2; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 7.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GULLALL 1140  
 DB 12 GULLALL 19

RESULT 21  
 AF0856  
 conserved hypothetical protein STY3056 [imported] - Salmonella enterica subsp. enterica  
 C/Species: Salmonella enterica subsp. enterica serovar Typhi  
 A/Note: this species has also been called Salmonella typhi  
 C/Date: 09-Nov-2001 #sequence, revision 09-Nov-2001 #text, change 27-Nov-2001  
 C/Accession: AF0856

R.Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Nature 413, 848-852, 2001  
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AF0856  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-103 <PAR>  
 A:Cross-References: GB:AL513382; PIDN:CAD06037.1; PID:g16504004; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY3056  
 C:Superfamily: hypothetical protein HI0673

Query Match 0.7%; Score 8; DB 2; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLALLLV 1141  
 |||||  
 6 LLLALLLV 13

RESULT 22  
 B75301  
 hypothetical protein - *Deinococcus radiodurans* (strain R1)

C:Species: *Deinococcus radiodurans*  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
 C:Accession: B75301  
 R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shih, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaluski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
 A:Reference number: A75250; PMID:20036896; PMID:10567266

A:Accession: B75301  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-127 <WHI>  
 A:Cross-References: GB:AE002054; GB:AE000513; NID:g6460010; PIDN:AAF11763.1; PID:g646001  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR2210  
 C:Superfamily: 1

A:Map position: 1  
 C:Superfamily: *Deinococcus radiodurans* hypothetical protein DR2210

Query Match 0.7%; Score 8; DB 2; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLALL 1140  
 |||||  
 7 GLLALL 14

RESULT 23  
 G82058  
 ribosomal protein L16 VC2589 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)  
 C:Species: *Vibrio cholerae*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: G82058  
 R.Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Basu, S.; Qin, H.; Dragol, I.; Sellers, F.  
 1; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A:Reference number: AB2035; PMID:20406833; PMID:10952301  
 A:Accession: G82058  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-136 <HEI>

A:Cross-References: GB:AE004326; GB:AE003852; NID:g9657161; PIDN:AAF95730.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC2589  
 A:Map position: 1  
 C:Superfamily: *Escherichia coli* ribosomal protein L16

Query Match 0.7%; Score 8; DB 2; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1055 LAKGTEVS 1062  
 |||||  
 20 LAKGTEVS 27

RESULT 24

S15785  
 heat-stable antigen-related hypothetical protein HSA-C - mouse

C:Species: *Mus musculus* (house mouse)  
 C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 05-Nov-1999  
 C:Accession: S15785  
 R.Wenger, R.H.; Ayane, M.; Bose, R.; Koehler, G.; Nielsen, P.J.  
 Eur. J. Immunol. 21, 1039-1046, 1991

A:Title: The genes for a mouse hematopoietic differentiation marker called the heat-stab  
 A:Reference number: S15783; PMID:91209380; PMID:2019286

A:Accession: S15785  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-141 <WEN>  
 A:Cross-References: EMBL:X56486; NID:g51441; PIDN:CAA39843.1; PID:g51442  
 A:Note: the authors translated the codon TTC for residue 87 as Ser and AGA for residue 1  
 C:Genetics:  
 A:Note: the authors did not translate the codon for residue 101

A:Introns: #status absent  
 C:Keywords: glycoprotein; membrane protein; phosphatidylinositol linkage

Query Match 0.7%; Score 8; DB 2; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLALL 1140  
 |||||  
 12 GLLALL 19

RESULT 25

AB1010  
 formylmethionine deformylase (EC 3.5.1.31) - *Salmonella enterica* subsp. *enterica* serovar

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh  
 A:Note: this species has also been called *Salmonella typhi*  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AB1010  
 R.Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
 A:Reference number: AB0502; PMID:11677608

A:Accession: AB1010  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-169 <PAR>  
 A:Cross-References: GB:AL513382; PIDN:CAD09179.1; PID:g16505183; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: fms  
 C:Superfamily: polypeptide deformylase  
 C:Keywords: hydrolase

Query Match 0.7%; Score 8; DB 2; Length 169;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 LVGKLFID 198  
|||||  
Db 138 LVGKLFID 145

RESULT 26  
H96977  
Probable membrane protein (imported) - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: H96977  
R:Willing, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Bacteriol. 183, 4823-4838, 2001  
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: H96977  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-185 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK76611.1; PID:G15023507; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
A:Genetics: CAC0634

Query Match 0.7%; Score 8; DB 2; Length 185;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 636 AAILSSR 643  
|||||  
Db 54 AAILSSR 61

RESULT 27  
S15102  
eosinophil major basic protein 2 precursor - guinea pig  
C:Species: Cavia porcellus (guinea pig)  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 21-Jul-2000  
C:Accession: S15102; S18501  
R:ROCK, I.; Shindoh, Y.; Nishida, T.; Nakai, S.; Hong, Y.M.; Mio, M.; Saito, T.; Tasaka, FBS Lett. 282, 56-60, 1991  
A>Title: Comparison of the amino acid and nucleotide sequences between human and two guinea pig eosinophil major basic proteins  
A:Reference number: S15102; MUID:91224343; PMID:2026266  
A:Accession: S15102  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-234 <NOX>  
A:Cross-references: DDBJ:DD0817; NID:G3135095; PIDN:BA00697.1; PID:G220293  
A:Accession: S18501  
A:Molecule type: protein  
A:Residues: 116, 'X', 118-134, 'X', 136-137, 'X', 139-145, 161-176, 181-200 <NOX>  
C:Superfamily: eosinophil major basic protein precursor; C-type lectin homology  
C:Keywords: antibiotic; chondroitin sulfate proteoglycan; cytolysin; eosinophil; glycoprotein  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-115/Domain: propeptide #status predicted <PRO>  
F:116-234/Product: eosinophil major basic protein 2 #status experimental <MAT>  
F:117-232/Domain: C-type lectin homology <LCH>  
F:24,25/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:69/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted  
F:135-232,209-224/Disulfide bonds: #status predicted

Query Match 0.7%; Score 8; DB 1; Length 234;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141  
|||||  
Db 4 LLLALLV 11

RESULT 28

E72629  
hypothetical protein APE1493 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
C:Accession: E72629  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: E72629  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-240 <KAW>  
A:Cross-references: DDBJ:AP000061; NID:G5104821; PIDN:BA080491.1; PID:dl044277; PID:G510  
A:Experimental source: strain K1  
A:Genetics: APE1493  
C:Superfamily: Aeropyrum pernix hypothetical protein APE1493

Query Match 0.7%; Score 8; DB 2; Length 240;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 GYSVSSML 439  
|||||  
Db 107 GYSVSSML 114

RESULT 29  
A81345  
histidine-binding protein precursor Cj0734c (imported) - Campylobacter jejuni (strain NC82)  
C:Species: Campylobacter jejuni  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: A81345  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chillingworth, R.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett, A.; et al. Nature 403, 665-668, 2000  
A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypoxanthine auxotrophy  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: A81345  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-251 <PAR>  
A:Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CA873008.1; PID:G696819  
A:Experimental source: serotype O2, strain NCTC 1168  
A:Genetics: Cj0734c  
C:Superfamily: lysine-arginine-ornithine-binding protein

Query Match 0.7%; Score 8; DB 2; Length 251;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 GTLGNTA 929  
|||||  
Db 146 GTLGNTA 153

RESULT 30  
B55904  
C 3.4.25.1 proteasome endopeptidase complex ( ) chain PRE6 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: proteasome alpha chain; proteasome PRE6; protein O2065; protein YOL03  
C:Species: Saccharomyces cerevisiae  
C>Date: 19-Jan-1996 #sequence\_revision 19-Jan-1996 #text\_change 03-Jun-2002  
C:Accession: B55904; S66721; S66723; S47910  
R:Heinemyer, W.; Treondle, N.; Albrecht, G.; Wolf, D.H. Biochemistry 33, 12229-12237, 1994  
A>Title: PRE5 and PRE6, the last missing genes encoding 20S proteasome subunits from yeast  
A:Reference number: B55904; MUID:95001940; PMID:7918444  
A:Accession: B55904  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-254 <HE1>  
 A:Cross-references: GB:L34348; NID:G511904; PIDN:AAA4903.1; PID:G511905  
 R:Habibig, B.; Hattenberg, U.; Hollenberg, C.P.; Ramezani Rad, M.  
 Submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S66703  
 A:Accession: S66721  
 A:Molecule type: DNA  
 A:Residues: 1-254 <HAB>  
 A:Cross-references: EMBL:Z74780; NID:G1419832; PIDN:CAA9040.1; PID:G1419833; MIPS:YOL03  
 A:Experimental source: strain S288C  
 R:Amstrong, W.; Bense, V.; Rechmann, S.; Schwager, C.; Teodoru, C.; Voss, H.; Wiemann, S.  
 Submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S66723  
 A:Accession: S66723  
 A:Molecule type: DNA  
 A:Residues: 1-254 <ANS>  
 A:Cross-references: EMBL:Z74780; NID:G1419832; PIDN:CAA9040.1; PID:G1419833; MIPS:YOL03  
 A:Experimental source: strain S288C  
 R:Heinmeyer, W.; Troendle, N.; Albrecht, G.; Wolf, D.H.  
 Submitted to the EMBL Data Library, July 1994  
 A:Description: PRE5 and PRE6, the last missing genes encoding 20S proteasome subunits fr  
 A:Reference number: S47908  
 A:Accession: S47910  
 A:Molecule type: DNA  
 A:Residues: 1-254 <HEM>  
 A:Cross-references: EMBL:U34348; NID:G511904; PIDN:AAA4903.1; PID:G511905  
 C:Genetics:  
 A:Gene: SGD:PRE6  
 A:Cross-references: SGD:S0005398; MIPS:YOL038w  
 A:Map position: 15L  
 C:Superfamily: multicatalytic endopeptidase complex chain C9  
 C:Keywords: hydrolase; proteinase

Query Match 0.7%; Score 8; DB 2; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1109 SLEEVQQT 1116  
 |||||  
 DB 198 SLEEVQQT 205

RESULT 31  
 T37985  
 proteasome component SPBC106.16 - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C:Accession: T37985; T39273  
 R:Murphy, U.; Harris, D.; Barrett, B.G.; Rajandream, M.A.; Walsh, S.V.  
 Submitted to the EMBL Data Library, March 1996  
 A:Reference number: Z21759  
 A:Accession: T37985  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-259 <MUR>  
 A:Cross-references: EMBL:Z69909; PIDN:CAA93796.1; GSPDB:GN00066; SPDB:SPAC19G10.14C  
 A:Experimental source: strain 972h-; cosmid c19G10  
 R:Lyme, M.; Rajandream, M.A.; Barrett, B.G.; Davis, P.; Churcher, C.M.  
 Submitted to the EMBL Data Library, August 1999  
 A:Reference number: Z21840  
 A:Accession: T39273  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-259 <LYN>  
 A:Cross-references: EMBL:AL110295; PIDN:CAE53732.1; GSPDB:GN00067; SPDB:SPBC106.16  
 A:Experimental source: strain 972h-; cosmid c106  
 C:Genetics: <MUR1>  
 A:Gene: SPAC19G10.14C  
 A:Map position: 1  
 A:Introns: 33/3; 158/2  
 C:Genetics: <LYN1>  
 A:Gene: SPBC106.16  
 A:Map position: 2

C:Superfamily: multicatalytic endopeptidase complex chain C9

Query Match 0.7%; Score 8; DB 2; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1109 SLEEVQQT 1116  
 |||||  
 DB 194 SLEEVQQT 201

RESULT 32  
 S22090  
 catechol O-methyltransferase (EC 2.1.1.6) - rat  
 N:Alternate names: catechol-O-methyltransferase  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jul-2000  
 C:Accession: S22090; JQ0787; I52356  
 R:Tenhunen, J.; Salminen, M.; Jalanko, A.; Ukkonen, S.; Uimannen, I.  
 Submitted to the EMBL Data Library, June 1992  
 A:Description: Structure of the rat catechol-O-methyltransferase gene: Separate promoter  
 A:Reference number: S22090  
 A:Accession: S22090  
 A:Molecule type: DNA  
 A:Residues: 1-264 <TEN>  
 A:Cross-references: EMBL:Z12651; NID:G55889; PIDN:CAA78276.1; PID:G55890  
 R:Salminen, M.; Lundstroem, K.; Tilgmann, C.; Savolainen, R.; Kalkkinnen, N.; Uimannen, I.  
 Gene 93, 241-247, 1990  
 A>Title: Molecular cloning and characterization of rat liver catechol-O-methyltransferas  
 A:Reference number: JQ0787; MUID:91035034; PMID:2227437  
 A:Accession: JQ0787  
 A:Molecule type: mRNA  
 A:Residues: 44-264 <SAL>  
 A:Cross-references: GB:M60753; NID:G203336; PIDN:AAA40881.1; PID:G203337  
 A:Experimental source: liver  
 R:Tenhunen, J.; Uimannen, I.  
 Biochem. J. 296, 595-600, 1993  
 A>Title: Production of rat soluble and membrane-bound catechol O-methyltransferase forms  
 A:Reference number: I52356; MUID:94107221; PMID:8280056  
 A:Accession: I52356  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-264 <RES>  
 A:Cross-references: EMBL:Z12651; NID:G55889; PIDN:CAA78276.1; PID:G55890  
 C:Comment: This enzyme catalyzes the transfer of a methyl group from S-adenosyl-L-methio  
 C:Genetics:  
 A:Gene: COMT  
 C:Superfamily: caffeineyl-CoA 3-O-methyltransferase  
 C:Keywords: methyltransferase; S-adenosylmethionine; transmembrane protein

Query Match 0.7%; Score 8; DB 2; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLLLALL 1140  
 |||||  
 DB 9 GLLLLALL 16

RESULT 33  
 S18541  
 hypothetical protein 3 - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 22-Oct-1999  
 C:Accession: S18541  
 R:Caillero, J.L.; Martinez, E.; Malpartida, F.; Hopwood, D.A.  
 Mol. Gen. Genet. 230, 401-412, 1991  
 A>Title: Organisation and functions of the activa region of the actinorhodin biosynthetic  
 A:Reference number: S18539; MUID:92114870; PMID:1766437  
 A:Accession: S18541  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-281 <CAB>

A:Cross-references: EMBL:X58833; NID:g46812; PIDN:CAA41639.1; PID:g46815

Query Match 0.7%; Score 8; DB 2; Length 281;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 708 TAGARAAAF 715

DB 89 TAGARAAAF 96

RESULT 34

F71556 probable had hydrolase/phosphatase - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C:Species: Chlamydia trachomatis

C>Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999

C:Accession: F71556

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis

A:Reference number: A71570; MUID:99000809; PMID:9784136

A:Accession: F71556

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <ARN>

A:Cross-references: GB:AE001284; GB:AE001273; NID:g3328494; PIDN:AAC67694.1; PID:g332850

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: CT103

Query Match 0.7%; Score 8; DB 2; Length 288;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 PEARLTV 272

DB 141 PEARLTV 148

RESULT 35

C65217 alliose kinase (EC 2.7.1.55) - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: C65217

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A:Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: C65217

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-309 <BLAT>

A:Cross-references: GB:AE000482; GB:U00096; NID:g2367349; PIDN:AACT7045.1; PID:g1790522;

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C:Accession: A85016

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617158

A:Accession: A85016

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-360 <STO>

A:Cross-references: GB:NC\_001268; NID:g7267619; PIDN:CAB80931.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4G01220

A:Map position: 4

Query Match 0.7%; Score 8; DB 2; Length 360;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141

DB 44 LLLALLLV 51

RESULT 37

S19172 cytochrome P450 2B4 - rat (fragments)

N:Alternate names: cytochrome P450 LM2

N:Comments: oxidoreductase (EC 1.-.-.-)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 22-Nov-1993 #sequence\_revision 21-Jul-1995 #text\_change 04-Mar-2000

C:Accession: S19172

R:Yuan, P.M.; Ryan, D.E.; Levin, W.; Shively, J.E.

Proc. Natl. Acad. Sci. U.S.A. 80, 1169-1173, 1983

A:Title: Identification and localization of amino acid substitutions between two phenols

A:Reference number: S19172; MUID:83144040; PMID:6572377

A:Accession: S19172

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-158;159-200;201-310;311-367 <YU>

A:Experimental source: strain Long-Evans

C:Genetics:

A:Gene: CYP2B4

C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology

C:Keywords: chromoprotein; heme; iron; metalloprotein; microsome; monooxygenase; oxidore-

F312/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 0.7%; Score 8; DB 2; Length 367;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141

DB 6 LLLALLLV 13

RESULT 38

T31771 hypothetical protein F59D6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T31771

R:Bradshaw, H.; Graves, T.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid F59D6.

A:Reference number: Z21083

A:Accession: T31771

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-383 <BRA>

A:Cross-references: EMBL:AF016435; PIDN:AAB65877.1; GSPDB:GN000023; CESP:F59D6.2

A:Experimental source: strain Bristol N2; clone F59D6

C:Genetics:  
A:Gene: CESP:FS9D6.2  
A:Map position: 5  
A:Introns: 72/1; 100/1; 262/3  
C:Superfamily: pepsin

Query Match 0.7%; Score 8; DB 2; Length 383;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1136 LLLALLVFC 1143  
DB 8 LLLALLVFC 15

RESULT 39  
G83568  
probable permease of ABC transporter PA0605 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83568  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: G83568  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-415 <STO-  
A:Cross-references: GB:AE004497; GB:AE004091; NID:g9946478; PIDN:AA03994.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0605

Query Match 0.7%; Score 8; DB 2; Length 415;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLLA 1138  
DB 388 LGGLLLLA 395

RESULT 40  
AC2275  
hypothetical protein alr3754 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AC2275  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2275  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-433 <KUR-  
A:Cross-references: GB:BA000019; PIDN:BA075453.1; PID:g17132888; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr3754

Query Match 0.7%; Score 8; DB 2; Length 433;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 982 VSGLIISA 989  
DB 247 VSGLIISA 254

RESULT 41  
T01721  
hypothetical protein A\_IG002N01.8 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 24-Mar-1999  
C:Accession: T01721  
R:Schneet, P.; Maggi, L.  
submitted to the EMBL Data Library, June 1997  
A:Description: The sequence of A. thaliana IG002N01.  
A:Reference number: Z14407  
A:Accession: T01721  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-444 <SCH>  
A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191131  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 68/1; 235/3; 302/3; 374/3  
A:Note: A\_IG002N01.8

Query Match 0.7%; Score 8; DB 2; Length 444;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141  
DB 115 LLLALLV 122

RESULT 42  
O4RTP  
cytochrome P450 2B1 - rat  
N:Alternate names: cytochrome P450 b; cytochrome P450, phenobarbital-inducible  
N:Contains: unspecified monooxygenase (EC 1.14.14.1)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 18-Aug-1982 #sequence\_revision 17-May-1996 #text\_change 03-Mar-2000  
C:Accession: A00176; A54251; A22363; A29298; S03854; A02255; I54796  
R:Fujii-Kuriyama, Y.; Mizukami, Y.; Kawajiri, K.; Sogawa, K.; Muramatsu, M.  
Proc. Natl. Acad. Sci. U.S.A. 79, 2793-2797, 1982  
A:Title: Primary structure of a cytochrome P-450: coding nucleotide sequence of phenobar  
A:Reference number: A93912; MUID:82222224; PMID:6953431  
A:Accession: A00176  
A:Molecule type: mRNA  
A:Residues: 6'491 <FU>  
A:Cross-references: EMBL:J00719; NID:G203752; PIDN:AAA1024.1; PID:G203753  
A:Note: The authors translated the codon GAT for residue 166 as Glu, CTG for residue 292  
R:Fujii-Kuriyama, Y.; Mizukami, Y.; Kawajiri, K.; Sogawa, K.; Muramatsu, M.  
Proc. Natl. Acad. Sci. U.S.A. 79, 5443, 1982  
A:Title: Primary structure of a cytochrome P450: coding nucleotide sequence of phenobar  
A:Reference number: A93925  
A:Contents: annotation  
A:Note: The translations shown in reference A93912 are acknowledged  
R:Roberts, E.S.; Hopkins, N.E.; Zaluzec, E.U.; Gage, D.A.; Alworth, W.L.; Hollenberg, P.  
Biochemistry 33, 3766-3771, 1994  
A:Title: Identification of active-site peptides from (3)H-labeled 2-ethylnaphthalene-i  
A:Reference number: A54251; MUID:94190899; PMID:8142377  
A:Accession: A54251  
A:Molecule type: protein  
A:Residues: 290-301, 'X' <ROB>  
R:Suwa, Y.; Mizukami, Y.; Sogawa, K.; Fujii-Kuriyama, Y.  
J. Biol. Chem. 260, 7980-7984, 1985  
A:Title: Gene structure of a major form of phenobarbital-inducible cytochrome P-450 in r  
A:Reference number: A22363; MUID:85234490; PMID:2989270  
A:Accession: A22363  
A:Molecule type: DNA  
A:Residues: 1-91, 'P', '93-204', 'V', '329-356', 'H', '358-391', 'R', '393-415', 'V', '417-433',  
A:Cross-references: GB:L00320; NID:g203816; PIDN:AAA1046.1; PID:g203818  
A:Note: The authors translated the codon CAG for residue 57 as Gly, CCT for residue 92 a  
as Glu, AAA for residue 236 as Leu, AGC for residue 259 as Asn, GTT for residue 328 as I  
as Arg

R.Rangarajan, P.N.; Ravishankar, H.; Padmanaban, G.  
 Biochem. Biophys. Res. Commun. 144, 258-263, 1987  
 A>Title: Isolation of a cytochrome P-450 gene variant and characterization of its 5' fl  
 A:Reference number: A29298; MUID:87213174; PMID:3579906  
 A:Accession: A29298  
 A>Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-57 <RAN>  
 R.Oesch, F.; Waxman, D.J.; Morrissey, J.J.; Honscha, W.; Kissel, W.; Friedberg, T.  
 Arch. Biochem. Biophys. 270, 23-32, 1989  
 A>Title: Antibodies targeted against hypervariable and constant regions of cytochromes P  
 A:Reference number: S03854; MUID:89192373; PMID:2539047  
 A:Accession: S03854  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-18,146-160, 'E', 162-165,166,330-361,362-380,402-423 <OES>  
 R.Botelho, L.H.; Ryan, D.E.; Levin, W.  
 J. Biol. Chem. 254, 5635-5640, 1979  
 A>Title: Amino acid compositions and partial amino acid sequences of three highly purified  
 or 3-methylcholanthrene.  
 A:Reference number: A92255; MUID:79194111; PMID:109438  
 A:Accession: A92255  
 A:Molecule type: protein  
 A:Residues: 1-3, 'T', 5-22 <BOT>  
 R.Fujii-Kuriyama, Y.; Mizukami, Y.; Taniguchi, T.; Muramatsu, M.  
 Int. Symp. Princess Takamatsu Cancer Res. Fund 12, 31-40, 1982  
 A>Title: Molecular cloning and coding nucleotide sequence of complementary DNA of cytoch  
 A:Reference number: I54796; MUID:83160754; PMID:6300027  
 A:Accession: I54796  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 6-491 <RES>  
 A:Cross-references: GB:M37134; NID:G203784; PIDN:AAC2028.1; PID:G203785  
 C:Genetics:  
 A:Gene: CYP2B1  
 A:Introns: 57/3; 112/1; 162/1; 215/3; 274/3; 322/1; 384/3; 432/1  
 C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
 C:Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo  
 F:295-458/Domain: cytochrome P450 homology <P45>  
 F:302/Active site: Thr #status predicted  
 F:336/Binding site: heme iron (Cys) (axial ligand) #status predicted  
 Query Match 0.7%; Score 8; DB 1; Length 491;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141  
 |||||  
 DB 6 LLLALLV 13

RESULT 43  
 OARTP2  
 cytochrome P450 2B2 - rat  
 N:Alternate names: cytochrome P450 PB-4; cytochrome P450, phenobarbital-inducible; cyto  
 N:Contains: oxidoreductase (EC 1.-.-.-)  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 04-Dec-1986 #sequence revision 17-May-1996 #text change 01-Dec-2000  
 C:Accession: A21162; A00177; B00176; B92255; S15589; A21872; A27236; S03855; I59060  
 R.Mizukami, Y.; Sogawa, K.; Suwa, Y.; Muramatsu, M.; Fujii-Kuriyama, Y.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 3958-3962, 1983  
 A>Title: Gene structure of a phenobarbital-inducible cytochrome P-450 in rat liver.  
 A:Reference number: A21162; MUID:83247397; PMID:6306654  
 A:Accession: A21162  
 A:Molecule type: DNA  
 A:Residues: 1-472, 'W', 474-491 <MT2>  
 A:Cross-references: EMBL:000728; NID:G203845; PIDN:AAA41056.1; PID:G203847  
 A>Note: the authors translated the codon AGR for residue 4 as Thr, and ATG for residue 3  
 J.Frey, A.B.; Waxman, D.J.; Kreibich, G.  
 J. Biol. Chem. 260, 15253-15255, 1985  
 A>Title: The structure of phenobarbital-inducible rat liver cytochrome P-450 isoenzyme F  
 A:Reference number: A00177; MUID:86059379; PMID:3877725  
 A:Accession: A00177

A:Molecule type: protein  
 A:Residues: 1-291, 'P', 293-320, 'AE', 323-475, 'D', 477-491 <FRE>  
 R.Fujii-Kuriyama, Y.; Mizukami, Y.; Kawajiri, K.; Sogawa, K.; Muramatsu, M.  
 Proc. Natl. Acad. Sci. U.S.A. 79, 2793-2797, 1982  
 A>Title: Primary structure of a cytochrome P-450: coding nucleotide sequence of phenobar  
 A:Reference number: A93912; MUID:82222224; PMID:695431  
 A:Accession: B00176  
 A:Molecule type: mRNA  
 A:Residues: 6-359, 'S', 361-362, 'V', 364-366, 'V', 368-406, 'S', 408-416, 'N', 418, 'A', 420-477, 'G  
 A>Note: nucleotide sequence for residues 1-5 is not given  
 A>Note: the authors translated the codon CAT for residue 166 as Glu, CTG for residue 292  
 R.Fujii-Kuriyama, Y.; Mizukami, Y.; Kawajiri, K.; Sogawa, K.; Muramatsu, M.  
 Proc. Natl. Acad. Sci. U.S.A. 79, 5443, 1982  
 A>Title: Primary structure of a cytochrome P450: coding nucleotide sequence of phenobarb  
 A:Reference number: A93925  
 A:Contents: annotation; revisions  
 A>Note: the mistranslations in reference A93912 are acknowledged  
 R.Botelho, L.H.; Ryan, D.E.; Levin, W.  
 J. Biol. Chem. 254, 5635-5640, 1979  
 A>Title: Amino acid compositions and partial amino acid sequences of three highly purified  
 or 3-methylcholanthrene.  
 A:Reference number: A92255; MUID:79194111; PMID:109438  
 A:Accession: B92255  
 A:Molecule type: protein  
 A:Residues: 1-3, 'T', 5-22 <BOT>  
 R.Lacroix, D.; Desrochers, M.; Lambert, M.; Anderson, A.  
 Gene 86, 201-207, 1990  
 A>Title: Alternative splicing of mRNA encoding rat liver cytochrome P450e (P450IIB2).  
 A:Reference number: S15589; MUID:90215299; PMID:2325573  
 A:Accession: S15589  
 A:Molecule type: mRNA  
 A:Residues: 105-113, 'P', 115-274, 'VSPAMRE', 275-321, 'E', 323-491 <LAC>  
 A:Cross-references: EMBL:M34452; NID:G203679; PIDN:AAA41004.1; PID:G203680  
 A>Note: translation of the nucleotide sequence is not complete  
 R.Philips, I.R.; Shephard, E.A.; Ashworth, A.; Rabin, B.R.  
 Gene 24, 41-52, 1983  
 A:Reference number: A21872  
 A:Accession: A21872  
 A:Molecule type: mRNA  
 A:Residues: 168-321, 'E', 323-443, 'K', 445-491 <PHI>  
 R.Affolter, M.; Anderson, A.  
 Biochem. Biophys. Res. Commun. 118, 655-662, 1984  
 A>Title: Segmental homologies in the coding and 3' non-coding sequences of rat liver cyt  
 A:Reference number: A32736; MUID:84153837; PMID:632758  
 A:Accession: A32736  
 A:Molecule type: mRNA  
 A:Residues: 385-491 <AFP>  
 A:Cross-references: GB:K01626; NID:G203782; PIDN:AAA41037.1; PID:G203783  
 R.Oesch, F.; Waxman, D.J.; Morrissey, J.J.; Honscha, W.; Kissel, W.; Friedberg, T.  
 Arch. Biochem. Biophys. 270, 23-32, 1989  
 A>Title: Antibodies targeted against hypervariable and constant regions of cytochromes P  
 A:Reference number: S03854; MUID:89192373; PMID:2539047  
 A:Accession: S03854  
 A>Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 323-431 <RES>  
 A:Cross-references: GB:M3234; NID:G203848; PIDN:AAA41057.1; PID:G554434  
 C:Genetics:  
 A:Gene: CYP2B2  
 A:Introns: 384/3  
 C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
 C:Keywords: alternative splicing; chromoprotein; electron transfer; endoplasmic reticul  
 F:295-458/Domain: cytochrome P450 homology <P45>  
 F:436/Binding site: heme iron (Cys) (axial ligand) #status predicted



```

Query Match      0.7%; Score 8; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 32;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1134 LLLALLLV 1141
        |||||
Db      6 LLLALLLV 13

RESULT 44
B31047
testosterone 16alpha-hydroxylase (EC 1.14.14.-) cytochrome P450 2B10 - mouse
N/Alternate names: cytochrome P450CB1; cytochrome P450p3/46
C/Species: Mus musculus (house mouse)
C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 04-Mar-2000
C/Accession: B31047; A60559
R/NoShiro, M.; Lakeo, M.; Kawajiri, K.; Negishi, M.
Biochemistry 27, 6434-6443, 1988
A/Title: R1P locus: regulation of female-specific isozyme (I-P-450-16-alpha) of testosterone
A/Reference number: A31047; MUID:89118235; PMID:3219345
A/Accession: B31047
A/Molecule type: mRNA
A/Residues: 1-500 <NO2>
A/Cross-references: EMBL:M21856; NID:g201968; PIDN:AAA40425.1; PID:g201969
A/Note: the authors translated the codon TTA for residue 55 as Phe, AGA for residue 133 as Val, AAA for residue 236 as Asn, AGG for residue 251 as Lys, AAG for residue 253 as 3 as Asp
A/Note: the authors translated the codon GCA for residue 281 as Thr, TTC for residue 283 as Arg, ACA for residue 321 as Ala, CTG for residue 331 as Leu, TCA for residue 349 as e Ala, and GAC for residue 479 as Gly
A/Note: the sequence nucleotide translation from Fig. 8 is inconsistent with the nucleot R/Bornheim, L.M.; Correia, M.A.
Mol. Pharmacol. 36, 377-383, 1989
A/Title: Purification and characterization of a mouse liver cytochrome P-450 induced by A/Reference number: A60559; MUID:89384475; PMID:2779523
A/Accession: A60559
A/Molecule type: protein
A/Residues: 1-15 <BOR>
A/Note: this enzyme was induced by cannabidiol
C/Genetics:
A/Gene: CYP2b-10
C/Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C/Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo F/295-467/Domain: cytochrome P450 homology <P45>
F/445/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match      0.7%; Score 8; DB 2; Length 500;
Best Local Similarity 100.0%; Pred. No. 33;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1134 LLLALLLV 1141
        |||||
Db      6 LLLALLLV 13

RESULT 45
139311
Kruppel-type zinc finger protein ZNF74 - human
C/Species: Homo sapiens (man)
C/Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 01-Dec-2000
C/Accession: 139311; F42825; S37482
R/Aubry, M.; Demczuk, S.; Desmazet, C.; Alkem, M.; Aurias, A.; Julien, J.P.; Rouleau, G.A Hum. Mol. Genet. 2, 1583-1587, 1993
A/Title: Isolation of a zinc finger gene consistently deleted in Digorge syndrome.
A/Reference number: 139311; MUID:94093543; PMID:8268910
A/Accession: 139311
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-572 <RES>
A/Cross-references: EMBL:X71623; NID:G1050529; PIDN:CAA50632.1; PID:G1050530
A/Note: submitted to the EMBL Data Library, April 1993, revised 01-NOV-1995
R/Aubry, M.; Maineau, C.; Zhang, F.R.; Zahed, L.; Figlewicz, D.; Delattre, O.; Thomas, Genomics 13, 641-648, 1992

```

```

A/Title: Cloning of six new genes with zinc finger motifs mapping to short and long arms
A/Reference number: A42825; MUID:92347859; PMID:1639391
A/Accession: F42825
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: nucleic acid
A/Residues: 427-444, 'V', 446-511 <AUB>
A/Note: sequence extracted from NCBI backbone (NCBI:109775)
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match      0.7%; Score 8; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 37;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      898 FSCSLLS 905
        |||||
Db      269 FSCSLLS 276

RESULT 46
136917
glycoprotein IIB - yellow baboon (fragment)
C/Species: Papio cynocephalus, Papio hamadryas cynocephalus (yellow baboon)
C/Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 29-Sep-1999
C/Accession: 136917
R/Hayzer, D.O.; Shoji, M.; Kim, T.M.; Runge, M.S.; Hanson, S.R.
Gene 151, 267-271, 1994
A/Title: Alternative splicing of the mRNA encoding baboon glycoprotein receptor GPIIB.
A/Reference number: 136916; MUID:95129873; PMID:782888
A/Accession: 136917
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-604 <RES>
A/Cross-references: GB:L12233; NID:G176584; PIDN:AAA65936.1; PID:G176585
C/Superfamily: integrin alpha-2b chain

Query Match      0.7%; Score 8; DB 2; Length 604;
Best Local Similarity 100.0%; Pred. No. 38;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1130 VLGGLLLL 1137
        |||||
Db      569 VLGGLLLL 576

RESULT 47
S60658
legumain - Gnetum gnemon
N/Alternate names: IIS globulin; seed storage protein
C/Species: Gnetum gnemon
C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Aug-1999
C/Accession: S60658
R/Braun, H.; Horstmann, C.; Baunlein, H.
submitted to the EMBL Data Library, August 1995
A/Description: Legumins of the Gnetatae: characterization and evolutionary relationships
A/Reference number: S60658
A/Accession: S60658
A/Molecule type: mRNA
A/Residues: 1-607 <BRA>
A/Cross-references: EMBL:Z50779; NID:G949870; PIDN:CAA00642.1; PID:G949871
C/Superfamily: glycinin
C/Keywords: seed; storage protein

Query Match      0.7%; Score 8; DB 2; Length 607;
Best Local Similarity 100.0%; Pred. No. 39;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1134 LLLALLLV 1141
        |||||
Db      10 LLLALLLV 17

RESULT 48
A69845

```

Na+/H+ antiporter homolog yjbQ - *Bacillus subtilis*  
 C:Species: *Bacillus subtilis*  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C/Accession: A69845  
 R:Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet  
 C.; Bron, S.; Broullet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
 A.; Erlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier  
 A.; Lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koeth, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueid  
 Y, M.; Osawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelg  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, S.; Sadate, Y.; Sato, T.; Scanlon,  
 A.; Authors: Schlicht, S.; Schreier, R.; Scoffone, F.; Seguchi, J.; Sekowska, A.; Serot  
 A.; Whetters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yamamoto, K.; Yata, K.; Yoshida, K  
 A.; Authors: Yoshikawa, H.F.; Zmistein, E.; Yoshikawa, H.; Danchin, A.  
 A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A/Reference number: A69580; MUID:98044033; PMID:9384377  
 A/Accession: A69845  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-614 <KUN>  
 A/Cross-references: GB:Z99110; GB:AL009126; NID:92633472; PIDN:CAB13021.1; PID:el183184;  
 A/Experimental source: strain 168  
 C/Genetics:  
 A:Gene: yjbQ

Query Match 0.7%; Score 8; DB 2; Length 614;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141  
 |||||  
 DB 313 LLLALLV 320

RESULT 49  
 AC0527  
 ferrichrome protein protein FhuB precursor STY0221 [imported] - *Salmonella enterica* su  
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh  
 A/Note: this species has also been called *Salmonella typhi*  
 C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C/Accession: AC0527  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 Th. T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
 A>Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero  
 A/Reference number: AB0502; PMID:11677608  
 A/Accession: AC0527  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-685 <PAR>  
 A/Cross-references: GB:AL513382; PIDN:CAD01355.1; PID:gl6501482; GSPDB:GN00176  
 C/Genetics:  
 A:Gene: STY0221  
 C/Superfamily: vitamin B12 transport protein btuC

Query Match 0.7%; Score 8; DB 2; Length 685;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141  
 |||||  
 DB 36 LLLALLV 43

RESULT 50  
 G96661  
 hypothetical protein F24D7.13 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C/Accession: G96661  
 R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzman, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A/Reference number: A66141; MUID:21016719; PMID:11130712  
 A/Accession: G96661  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-767 <STO>  
 A/Cross-references: GB:AE005173; NID:96456164; PIDN:AA09152.1; GSPDB:GN00141  
 C/Genetics:  
 A:Gene: F24D7.13  
 A/Map position: 1

Query Match 0.7%; Score 8; DB 2; Length 767;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 AVGAPLED 575  
 |||||  
 DB 559 AVGAPLED 566

RESULT 51  
 EB3349  
 probable ClpA/B-type proteinase PA2371 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
 C:Species: *Pseudomonas aeruginosa*  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C/Accession: EB3349  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Waresner, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
 A/Reference number: AB2950; MUID:20437337; PMID:10984043  
 A/Accession: EB3349  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-849 <STO>  
 A/Cross-references: GB:AE004663; GB:AE004091; NID:99948405; PIDN:AA05759.1; GSPDB:GN001  
 A/Experimental source: strain PA01  
 C/Genetics:  
 A:Gene: PA2371  
 C/Superfamily: endopeptidase Clp ATP-binding chain

Query Match 0.7%; Score 8; DB 2; Length 849;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALLL 1140  
 |||||  
 DB 111 GLLALLL 118

RESULT 52  
 J01674  
 protein kinase TMK1 (EC 2.7.1.-), receptor type precursor - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 11-Jun-1999  
 C/Accession: J01674  
 R:Chang, C.; Schaller, G.E.; Patterson, S.E.; Kwok, S.F.; Meyrowitz, E.M.; Blecker, A.  
 Plant Cell 4, 1263-1271, 1992  
 A>Title: The TMK1 gene from *Arabidopsis* codes for a protein with structural and biochemi  
 A/Reference number: J01674; MUID:93076110; PMID:1332795

A:Accession: J01674  
 A:Molecule type: DNA  
 A:Residues: 1-942 <CHA>  
 A:Cross-references: GB:L00670; NID:G16687; PIDN:AAA2876.1; PID:G16688  
 C:Superfamily: protein kinase Xa1; leucine-rich alpha-2-glycoprotein repeat homology; F  
 C:Keywords: ATP; autophosphorylation; glycoprotein; phosphotransferase; receptor; serine  
 F:11-22/Domain: signal sequence #status predicted <Sto>  
 F:23-942/Product: protein kinase TMX1 #status predicted <MRT>  
 F:65-88/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>  
 F:89-111/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>  
 F:112-160/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>  
 F:113-160/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>  
 F:161-186/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>  
 F:187-203/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>  
 F:210-223/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>  
 F:233-255/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>  
 F:256-278/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>  
 F:280-299/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>  
 F:300-323/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>  
 F:324-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>  
 F:363-386/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>  
 F:387-410/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>  
 F:411-434/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>  
 F:480-503/Domain: transmembrane #status predicted <TMH>  
 F:586-872/Domain: protein kinase #status predicted <KIN>  
 F:594-602/Region: protein kinase ATP-binding motif  
 F:86,99,158,164,171,363,533,587/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:616,634,717,719/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 0.7%; Score 8; DB 1; Length 942;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 8; Conservative 0; Indels 0; Gaps 0;

Qy 1128 GSVGLGL 1135  
 Db 486 GSVGLGL 493

RESULT 53  
 B60017  
 outer capsid protein VP2 - bluetongue virus (serotype 3, strain South Africa-VACC)  
 C:Species: bluetongue virus  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 16-Jul-1999  
 C:Accession: B60017  
 R:Gould, A.R.; Pritchard, L.I.  
 Virus Res. 17, 31-52, 1990  
 A:Title: Relationships amongst bluetongue viruses revealed by comparisons of capsid and  
 A:Reference number: A60017; MUID:91021485; PMID:2171239  
 A:Accession: B60017  
 A:Molecule type: genomic RNA  
 A:Residues: 1-959 <GOU>  
 A:Cross-references: GB:X55801; NID:9297130; PIDN:CAA39323.1; PID:9297131  
 C:Genetics:  
 A:Map position: segment 2  
 C:Superfamily: bluetongue virus VP2 protein  
 C:Keywords: capsid protein; glycoprotein  
 F:749,910/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.7%; Score 8; DB 1; Length 959;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 8; Conservative 0; Indels 0; Gaps 0;

Qy 1159 EKREKLE 1166  
 Db 798 EKREKLE 805

RESULT 54  
 T15446  
 hypothetical protein C07G1.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T15446

R:Hawkins, J.  
 submitted to the EMBL Data Library, May 1996  
 A:Description: The sequence of C. elegans cosmid C07G1.  
 A:Reference number: Z18352

A:Accession: T15446  
 A:Status: preliminary; translated from GB/EMBL/DBU  
 A:Molecule type: DNA  
 A:Residues: 1-969 <HAM>  
 A:Cross-references: EMBL:U58751; NID:q1326379; PID:q1326381; PIDN:AB00657.1; GSPDB:GN00  
 A:Experimental source: strain Bristol N2; clone C07G1  
 C:Genetics:  
 A:Gene: CESP.C07G1.4  
 A:Map position: 4  
 A:Introns: 110/3; 149/2; 515/1; 564/1; 600/2; 690/2; 755/3; 843/3; 884/3; 940/2

Query Match 0.7%; Score 8; DB 2; Length 969;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 8; Conservative 0; Indels 0; Gaps 0;

Qy 672 EAVCLTAA 679  
 Db 9 EAVCLTAA 16

RESULT 55  
 A60163  
 glycoprotein IIB - rat  
 C:Species: Rattus sp. (rat)  
 C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 13-Sep-1998  
 C:Accession: A60163; B60163  
 R:Poncz, M.; Newman, P.J.  
 Blood 75, 1282-1289, 1990  
 A:Title: Analysis of rodent platelet glycoprotein IIB: evidence for evolutionarily conse  
 A:Reference number: A60163; MUID:90181615; PMID:2310828

A:Accession: A60163  
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-1037 <PON>  
 A:Accession: B60163  
 A:Molecule type: protein  
 A:Residues: 32-50 <PON2>  
 C:Comment: This protein is proteolytically processed into a heavy chain and a light chain  
 C:Superfamily: integrin alpha-2b chain  
 C:Keywords: calcium; disulfide bond; glycoprotein

Query Match 0.7%; Score 8; DB 2; Length 1037;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 8; Conservative 0; Indels 0; Gaps 0;

Qy 1130 VLGGILLL 1137  
 Db 1003 VLGGILLL 1010

RESULT 56  
 A34269  
 integrin alpha-2b chain precursor - human  
 M:Alternate names: antigen CD41b; platelet adhesion receptor glycoprotein IIB; platelet  
 C:Species: Homo sapiens (man)  
 C:Date: 25-May-1990 #sequence\_revision 23-May-1997 #text\_change 29-Sep-1999  
 C:Accession: A34269; A29522; I57461; A28937; S00268; A28581; A36430; B32528; A32  
 R:Heidenreich, R.; Eisman, R.; Surrey, S.; Delgrosso, K.; Bennett, J.S.; Schwartz, E.; P  
 Biochemistry 29, 1232-1244, 1990  
 A:Title: Organization of the gene for platelet glycoprotein IIB.  
 A:Reference number: A34269; MUID:90212612; PMID:2322558  
 A:Accession: A34269  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-286, 'VP', 289-731, 'AI', 734-1039 <HEI>  
 A:Cross-references: GB:M33319; NID:g183506; GB:M3320; NID:g183506; GB:M34334; NID:g1835  
 A:Note: the authors translated the codon AAC for residue 819 as ARG, CGG for residue 931  
 A:Note: the codons given for 287-Gly (GTG), 288-Ala (CCC), and 733-Ala (ATC) are inconsi  
 R:Poncz, M.; Eisman, R.; Heidenreich, R.; Silver, S.M.; Vialatte, G.; Surrey, S.; Schwart

J. Biol. Chem. 262, 8476-8482, 1987  
A.Title: Structure of the platelet membrane glycoprotein IIb. Homology to the alpha subunit  
A.Reference number: A29522; PMID:87250457; PMID:2439501  
A.Accession: A29522  
A.Molecule type: mRNA  
A.Residues: 1-312, 'A', 314-327, 'H', 329-632, 'S', 634-1039 <PON>  
A.Cross-references: GB:J02764; NID:G190067; PIDN:AAA6014.1; PID:G190068  
A.Note: the codons given for residues 313-Gly (GCA) and 328-Asp (CAC) are inconsistent  
R:Frachet, P.; Uzan, G.; Thevenon, D.; Denarier, E.; Prandini, M.H.; Marguerite, G.  
Mol. Biol. Rep. 14, 27-33, 1990  
A.Title: GPIIb and GPIIb amino acid sequences deduced from human megakaryocyte cDNAs.  
A.Reference number: 157461; PMID:90265363; PMID:2345548  
A.Accession: 157461  
A.Status: preliminary; translated from GB/EMBL/DDJ  
A.Molecule type: mRNA  
A.Residues: 1-22, 'A', 24-312, 'A', 314-345, 'D', 347-564, 'D', 566-1028, 'HT', 1031-1039 <RES>  
A.Cross-references: GB:M34480; NID:G183510; PIDN:AAA5926.1; PID:G306794  
R:Prandini, M.H.; Denarier, E.; Frachet, P.; Uzan, G.; Marguerite, G.  
Biochem. Biophys. Res. Commun. 156, 595-601, 1988  
A.Title: Isolation of the human platelet glycoprotein IIb gene and characterization of cDNA  
A.Reference number: A28937; PMID:89025907; PMID:2845986  
A.Accession: A28937  
A.Molecule type: DNA  
A.Residues: 1-62, 1021-1028, 'HT', 1031-1039 <PRA>  
A.Cross-references: GB:M22568; NID:G183448; PIDN:AAA52587.1; PID:G553310; GB:M22569; NID  
R:Uzan, G.; Frachet, P.; Lajmanovich, A.; Prandini, M.H.; Denarier, E.; Duperray, A.; Le  
Eur. J. Biochem. 171, 87-93, 1988  
A.Title: CDNA clones for human platelet GPIIb corresponding to mRNA from megakaryocytes  
A.Reference number: S00268; PMID:88111709; PMID:3422188  
A.Accession: S00268  
A.Molecule type: mRNA  
A.Residues: 392-565, 'V', 567-728, 'E', 730-1039 <UZA>  
A.Cross-references: EMBL:X06831; NID:G35519; PIDN:CAA29987.1; PID:G35520  
A.Note: the authors translated the codon AGC for residue 24 as Tyr; the sequence from F  
R:Bray, P.F.; Rosa, J.P.; Johnston, G.I.; Shiu, D.T.; Cook, R.G.; Lau, C.; Kan, Y.W.; Mc  
J. Clin. Invest. 80, 1812-1817, 1987  
A.Title: Platelet glycoprotein IIb. Chromosomal localization and tissue expression.  
A.Reference number: A28581; PMID:88059639; PMID:3479442  
A.Accession: A28581  
A.Molecule type: mRNA  
A.Residues: 868-1039 <BR1>  
A.Cross-references: GB:M18085; NID:G183503; PIDN:AAA52597.1; PID:G183504  
R:Bray, P.F.; Leung, C.S.I.; Shuman, M.A.  
J. Biol. Chem. 265, 9587-9590, 1990  
A.Title: Human platelets and megakaryocytes contain alternately spliced glycoprotein IIb  
A.Reference number: A36430; PMID:90277633; PMID:2351656  
A.Accession: A36430  
A.Molecule type: protein  
A.Residues: 888-1039 <BRA2>  
A.Cross-references: GB:J05494  
A.Accession: B36430  
A.Molecule type: mRNA  
A.Residues: 888-947, 982-1039 <BRA3>  
A.Cross-references: GB:J05494  
A.Note: this form was found in platelets, megakaryocytes, and an erythroleukemia line, B  
R:Hizawa, A.; Matsukage, A.; Shiku, H.; Takahashi, T.; Naito, K.; Yamada, K.  
Blood 69, 560-564, 1987  
A.Title: Purification and partial amino acid sequence of human platelet membrane glycopr  
A.Reference number: A90731; PMID:87101510; PMID:3801670  
A.Accession: B32528  
A.Molecule type: protein  
A.Residues: 487-501 <HR1>  
A.Accession: A32528  
A.Molecule type: protein  
A.Residues: 'X', 1027, 'X', 1029, 'X', 1031-1038 <HR2>  
R:Calvete, J.J.; Henschen, A.; Gonzalez-Rodriguez, J.  
Biochem. J. 261, 561-568, 1989  
A.Title: Complete localization of the intrachain disulphide bonds and the N-glycosylatio  
A.Reference number: S05249; PMID:89374157; PMID:2775232  
A.Accession: S05249  
A.Molecule type: protein  
A.Residues: 32-41, 64-66, 73-79, 91-95, 'X', 97, 109-116, 123-129, 156-158, 160, 'X', 162-165, 172-1  
621, 667-690, 692-701, 709-710, 'X', 712-717, 'X', 725-730, 735-744 <CAL1>

A.Accession: S05254  
A.Molecule type: protein  
A.Residues: 903-910, 'X', 912, 1019-1024 <CAL2>  
R:Calvete, J.J.; Alvarez, M.V.; Rivas, G.; Hew, C.L.; Henschen, A.; Gonzalez-Rodriguez,  
Biochem. J. 261, 551-560, 1989  
A.Title: Interchain and intrachain disulphide bonds in human platelet glycoprotein IIb.  
A.Reference number: S05248; PMID:89374156; PMID:2476117  
A.Accession: S05248  
A.Molecule type: protein  
A.Residues: 903-922, 934-939 <CAL3>  
R:Calvete, J.J.; Schaefer, W.; Henschen, A.; Gonzalez-Rodriguez, J.  
FEBS Lett. 272, 37-40, 1990  
A.Title: Characterization of the beta-chain N-terminus heterogeneity and the alpha-chain  
A.Reference number: S12871; PMID:91032137; PMID:2226834  
A.Contents: annotation; chemical and mass spectrometric analyses  
R:Calvete, J.J.; Muniz-Diaz, E.  
FEBS Lett. 328, 30-34, 1993  
A.Title: Localization of an O-glycosylation site in the alpha-subunit of the human plate  
A.Reference number: S35059; PMID:93345693; PMID:7688323  
A.Accession: S35059  
A.Status: preliminary  
A.Molecule type: protein  
A.Residues: 735-745, 751-765, 766-774, 775-782, 783-795, 801-810, 811-820, 868-877, 'X', 879-881  
R:Calvete, J.J.; Fernandez, S.; Leung, L.L.; McGreggor, J.L.  
Biochem. J. 279, 419-425, 1991  
A.Title: Separation of important new platelet glycoproteins (GPIa, GPIC, GPIC\*, GPIIa an  
A.Reference number: A56793; PMID:92061944; PMID:1953640  
A.Accession: A56793  
A.Molecule type: protein  
A.Residues: 32-42 <CAT>  
A.Experimental source: platelet  
R:Loftus, J.C.; Plow, E.F.; Frelinger III, A.L.; D'Souza, S.E.; Dixon, D.; Lacy, J.; Sor  
Proc. Natl. Acad. Sci. U.S.A. 84, 7114-7118, 1987  
A.Title: Molecular cloning and chemical synthesis of a region of platelet glycoprotein I  
A.Reference number: A28411; PMID:88041058; PMID:2444974  
A.Accession: A28411  
A.Molecule type: mRNA  
A.Residues: 'REYDLCQVOMLPVPLVLEAKAGRSPEVRSRALPNP', 786-819, 'T', 821-853, 'A', 855-964, 'F  
A.Cross-references: GB:J02963; NID:G190069; PIDN:AAA60115.1; PID:G387027  
A.Experimental source: HEL erythroleukemia cell library  
A.Note: the 5' end of this mRNA contains Alu-like sequence shown by antibodies to synthe  
R:Nakogorenko, E.M.; Yakubenko, V.P.; Ingham, K.C.; Medved, L.V.  
Eur. J. Biochem. 237, 205-211, 1996  
A.Title: Thermal stability of individual domains in platelet glycoprotein IIbIIa.  
A.Reference number: S65437; PMID:96203926; PMID:8620874  
A.Accession: S65437  
A.Molecule type: protein  
A.Residues: 872 <MAK>  
A.Accession: S65438  
A.Molecule type: protein  
A.Residues: 32 <MKO>  
C.Comment: Platelet membrane glycoprotein IIb-IIa (CD41/CD61 heterodimer) is a receptor  
C.Genetics: GDB:ITGA2B; GP2B  
A.Cross-references: GDB:120012; OMIM:273800  
A.Map position: 17q21.32-17q21.32  
C.Keywords: integrin alpha-2b chain  
C.Keywords: alternative splicing; glycoprotein; heterodimer; platelet; pyroglyutamic acid  
F:1-31/Domain: signal sequence #status predicted <SIG>  
F:32-887/Product: platelet glycoprotein IIb alpha chain #status experimental <MAT>  
F:891-1039/Product: platelet glycoprotein IIb beta chain (form 1) #status experimental <  
F:903-1039/Product: platelet glycoprotein IIb beta chain (form 2) #status experimental <  
F:994-1018/Domain: transmembrane #status predicted <TM>  
F:1019-1039/Domain: intracellular #status predicted <CYT>  
F:46-280, 601, 963/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:87-96, 138-161, 177-198, 504-515, 521-576, 633-639, 705-718/Disulfide bonds: #status predict  
F:711/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:857-911, 916-921/Disulfide bonds: #status experimental  
F:878/Binding site: carbohydrate (Ser) (covalent) #status experimental  
F:891/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 0.7%; Score 8; DB 2; Length 1039;  
Best Local Similarity 100.0%; Pred. No. 60;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1130 VLGGLLL 1137  
 |||||  
 DB 1004 VLGGLLL 1011

RESULT 57  
 S40311  
 integrated - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #ext\_change 21-Jul-2000

C:Accession: S40311  
 R:Wehli, M.; D'Antonio, A.; Fearney, I.M.; Smith, R.J.; Wilcox, M.  
 Mech. Dev. 43, 21-36, 1993

A:Title: Cloning and characterization of alpha(P51), a novel *Drosophila melanogaster* int

A:Reference number: S40311; MUID:94059764; PMID:8240969

A:Accession: S40311  
 A:Status: preliminary

A:Molecule type: mRNA  
 A:Residues: 1-1146 <MEH>  
 A:Cross-references: EMBL:X73975; NID:g440143; PIDN:CAA52155.1; PID:g440144

C:Genetics:

A:Gene: FlyBase:mew  
 A:Cross-references: FlyBase:FBgn0004456

C:Superfamily: Integrin alpha-2b chain

Query Match 0.7%; Score 8; DB 2; Length 1146;  
 Best Local Similarity 100.0%; Pred. No. 65;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
 |||||

DB 1094 GGLLLAL 1101

RESULT 58

RHNU1B

cell surface glycoprotein CD11b precursor [validated] - human

N:Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac-1  
 leukocyte integrin alpha chain; neutrophil adherence receptor alphaM chain

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #ext\_change 31-Dec-2000

C:Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; 152567

R:Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.  
 J. Biol. Chem. 263, 12403-12411, 1988

A:Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD

B.  
 A:Reference number: A31108; MUID:88315033; PMID:2457584

A:Accession: A31108

A:Molecule type: mRNA

A:Residues: 1-1153 <COR>

A:Cross-references: GB:J03925; NID:g187284; PIDN:AAA59544.1; PID:g307148

A:Note: part of this sequence was confirmed by protein sequencing

R:Arnaut, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.  
 J. Cell Biol. 106, 2153-2158, 1988

A:Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M

A:Reference number: A28915; MUID:88257215; PMID:2454331

A:Accession: A28915

A:Molecule type: mRNA

A:Residues: 1-499,501-965, 'P', 967-1153 <ARN>

A:Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:g186935; PIDN:AAA594

A:Note: the authors translated the codon TAG for residue 1129 as Thr

A:Note: part of this sequence, including the amino end of the mature protein, was confir

R:Shelley, C.S.; Arnaut, M.A.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991

A:Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg

A:Reference number: A41600; MUID:92073318; PMID:1683702

A:Accession: A41600

A:Molecule type: DNA

A:Residues: 1-9 <SHS>

A:Cross-references: GB:M76724; NID:g180018; PIDN:AAA58410.1; PID:g553215

R:Arnaut, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.

Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
 A:Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion  
 A:Reference number: A34193; MUID:88190151; PMID:2833753

A:Accession: A30892

A:Molecule type: mRNA

A:Residues: 917-1042 <AR2>

A:Cross-references: GB:M18044  
 R:Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989

A:Title: cDNA sequence for the alphaM subunit of the human neutrophil adherence receptor

A:Reference number: A32218; MUID:89098893; PMID:2563162

A:Accession: A32218

A:Molecule type: mRNA

A:Residues: 9-1153 <HIC>  
 A:Cross-references: GB:J04145; NID:g189068; PIDN:AAA59903.1; PID:g386975

A:Note: part of this sequence was confirmed by protein sequencing

R:Pleming, J.C.; Pahl, H.L.; Gonzalez, D.N.; Smith, T.F.; Tenen, D.G.  
 J. Immunol. 150, 480-490, 1993

A:Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1r

n during evolution.

A:Reference number: A46526; MUID:93123748; PMID:8419480

A:Accession: A46526

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 9-1153 <ELE>

A:Cross-references: GB:S52227; NID:g263047; PIDN:AB24821.1; PID:g263049

A:Note: the last three bases of intron 13, CAG, are included in some but not all mature

R:Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaut, M.A.  
 Biochem. Biophys. Acta 874, 368-371, 1986

A:Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across si

A:Reference number: A50664; MUID:87076671; PMID:3539202

A:Accession: A26091

A:Molecule type: protein

A:Residues: 17-31 <PIE>

A:Experimental source: granulocytes

R:Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.  
 Blood 79, 865-870, 1992

A:Title: Characterization of the myeloid-specific CD11b promoter.

A:Reference number: 152567; MUID:92144986; PMID:1346576

A:Accession: 152567

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-9 <RES>

A:Cross-references: GB:M84477; NID:g180184; PIDN:AAA51960.1; PID:g553219

C:Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1

C:Genetics:

A:Gene: GDB:ITGM; CR3A

A:Cross-references: GDB:120599; OMIM:120980

A:Map position: 16p11.2-16p11.2

A:Note: promoter contains a GATA motif and two Sp1 consensus binding sites

C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo

C:Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>

F:17-1108/Domain: extracellular #status predicted <EXT>

F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>

F:465-472/Region: calcium/magnesium binding #status predicted

F:530-538/Region: calcium/magnesium binding #status predicted

F:593-601/Region: calcium/magnesium binding #status predicted

F:1109-1134/Domain: transmembrane #status predicted <TM>

F:1135-1153/Domain: intracellular #status predicted <INT>

F:86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding

Query Match 0.7%; Score 8; DB 1; Length 1153;  
 Best Local Similarity 100.0%; Pred. No. 66;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
 |||||

DB 1116 GGLLLAL 1123

## RESULT 59

RHWUIC

cell surface glycoprotein CD11c precursor - human  
N:Alternate names: leukocyte adhesion receptor p150,95 alpha chain

C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 22-Jun-1999

C:Accession: A36584; A35543; S00864  
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.

J. Biol. Chem. 265, 12750-12751, 1990  
A:Reference number: A36584

A:Contents: erratum

A:Accession: A36584

A:Molecule type: DNA

A:Residues: 1-1163 <COR>

A:Note: this revision to the sequence from reference A35543 includes the carboxyl end  
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.

J. Biol. Chem. 265, 2782-2788, 1990

A:Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.  
A:Reference number: A35543; MUID:90153906; PMID:2303426

A:Accession: A35543

A:Molecule type: DNA

A:Residues: 1-834 <CO2>

A:Note: this sequence has been revised in reference A36584  
R:Corbi, A.L.; Miller, L.D.; O'Connor, K.; Larson, R.S.; Springer, T.A.

EMBO J. 6, 4023-4028, 1987

A:Title: CDNA cloning and complete primary structure of the alpha subunit of a leukocyte  
A:Accession: S00864; MUID:8816645; PMID:3327687

A:Molecule type: mRNA

A:Residues: 1-755, 'L', 757-1163 <CO3>

A:Cross-references: GB:M81695; EMBL:Y00093; NID:9487829; PIDN:AA59180.1; PID:9487830

A:Note: part of this sequence was confirmed by protein sequencing

C:Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on my  
C:Genetics:

A:Gene: GDB:ITGAX; CD11C

A:Cross-references: GDB:119758; OMIM:151510

A:Map position: 16p11.2-16p11.2

C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homoc  
C:Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;  
F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>

F:20-1107/Domain: extracellular #status predicted <EXT>

F:149-319/Domain: von Willebrand factor type A repeat homology <VWA4>

F:1108-1133/Domain: transmembrane #status predicted <TM>

F:1134-1163/Domain: intracellular #status predicted <INT>  
F:61,89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status pre

Query Match 0.7%; Score 8; DB 1; Length 1163;

Best Local Similarity 100.0%; Pred. No. 66;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLALL 1139

DB 1115 GGLLLALL 1122

## RESULT 60

G84922

hypothetical protein At2g48050 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: G84922

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.L.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
Neus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84922

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1500 <STO>

A:Cross-references: GB:AE002093; NID:94249412; PIDN:AAD13709.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g48050

A:Map position: 2

Query Match 0.7%; Score 8; DB 2; Length 1500;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1065 LRLVHNE 1072

DB 1202 LRLVHNE 1209

## RESULT 61

I38414

transcription factor IIC, box B-binding chain - human

C:Species: Homo sapiens (man)  
C:Date: 16-Feb-1996 #sequence\_revision 16-Feb-1996 #text\_change 05-Nov-1999

C:Accession: I38414  
R:Leclerc, N.D.; Fahnestock, M.L.; Shen, Y.; Aebersold, R.; Berk, A.J.  
Proc. Natl. Acad. Sci. U.S.A. 91, 1652-1656, 1994

A:Title: Human TFIIC Box B-Binding Subunit.  
A:Reference number: A3132; MUID:94173888; PMID:8127861

A:Accession: I38414

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2109 <RES>

A:Cross-references: EMBL:U02619; NID:9414932; PIDN:AA17985.1; PID:9442362

C:Genetics:

A:Gene: GDB:GTF3C1; TFIIC

A:Cross-references: GDB:223512

A:Map position: 16p12.1-16p11.2

Query Match 0.7%; Score 8; DB 2; Length 2109;

Best Local Similarity 100.0%; Pred. No. 1,1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 REPARLL 271

DB 170 REPARLL 177

## RESULT 62

T13924

sdh protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000

C:Accession: T13924  
R:Nguyen, D.N.; Liu, Y.; Litkev, M.L.; Reinke, R.

submitted to the EMBL Data Library, February 1997

A:Description: Sidelick, a member of the immunoglobulin superfamily, is required for pat

A:Reference number: Z17809

A:Accession: T13924

A:Status: preliminary; translated from GB/EMBL/DBU

A:Molecule type: mRNA

A:Residues: 1-2222 <NCU>

A:Cross-references: EMBL:U88578; NID:94099554; PID:94099555; PIDN:AAD09632.1

C:Genetics:

A:Gene: sdh

A:Cross-references: FlyBase:FBgn0021764

Query Match 0.7%; Score 8; DB 2; Length 2222;

Best Local Similarity 100.0%; Pred. No. 1,1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 970 TTLRVONT 977

DB 1842 TTLRVONT 1849

## RESULT 63

CGH03A

collagen alpha 3 (VI) chain precursor [validated] - human

N.Contains: collagen alpha 3(VI) chain, splice form A9/N10(-)  
 C/Species: Homo sapiens (man)  
 C/Date: 21-Nov-1993 #sequence revision 12-Nov-1999 #ext change 15-Sep-2000  
 C/Accession: A59140; S13679; S24465; A57083; S28776; S00245; C31952; C29848; S26510; S48  
 R.Chu, M.L.  
 submitted to GenBank, May 1998  
 A/Reference number: A59140  
 A/Accession: A59140  
 A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 1-3176 <CHU>  
 A/Cross-references: GB:X52022; NID:G3127925; PIDN:CAA36267.1; PID:G3127926  
 R.Chu, M.L.; Zhang, R.Z.; Pan, T.; Stokes, D.; Conway, D.; Kuo, H.J.; Glanville, R.; May  
 EMBL J. 9, 385-393, 1990  
 A/Title: Mosaic structure of globular domains in the human type VI collagen alpha-3 chain  
 A/Reference number: S13679; MUID:905151612; PMID:1689238  
 A/Accession: S13679  
 A/Molecule type: mRNA  
 A/Residues: 1-30,237-313, 'CWV', 318-322, 'AR', 326-1815, 'PD', 1818-1819, 'ID', 1822-3176 <CHS>  
 A/Cross-references: EMBL:X52022; NID:G3127925  
 A/Accession: S24465  
 A/Molecule type: protein  
 A/Residues: 574-585,965-973, 'X', 975-976,1306-1325,1361-1377,1381-1401,1473-1506, 'X', 1508-1962, 'X', 1964-1965,2018-2037,2374-2410,2445-2459,2466-2469, 'X', 2471-2474,2504-2508, 'X',  
 R.Zanussi, S.; Dollana, R.; Segat, D.; Bonaldo, P.; Colombatti, A.  
 J. Biol. Chem. 267, 24082-24089, 1992  
 A/Title: The human type VI collagen gene. mRNA and protein variants of the alpha3 chain  
 A/Reference number: S28776; MUID:93054780; PMID:1339440  
 A/Accession: A57083  
 A/Molecule type: DNA  
 A/Residues: 310-328 <ZAN>  
 A/Accession: S28776  
 A/Molecule type: mRNA  
 A/Residues: 32-126, 'AK', 129-136, 'L', 138-236 <ZAZ>  
 A/Cross-references: GB:S49432; NID:G260296; PIDN:ABZ4261.1; PID:G260297  
 R.Chu, M.L.; Mann, K.; Deutzmann, R.; Pribula-Conway, D.; Hsu-Chen, C.C.; Bernard, M.P.;  
 Eur. J. Biochem. 168, 309-317, 1987  
 A/Title: Characterization of three constituent chains of collagen type VI by peptide seq  
 A/Reference number: S00245; MUID:88029444; PMID:366527  
 A/Accession: S00245  
 A/Molecule type: mRNA, protein  
 A/Residues: 2024-2046,2092-2156, 'R', 2203-2208, 'X', 2210-2211, 'X', 2213-2227,2228-2251,2314  
 A/Cross-references: GB:X06196; NID:G30055; PIDN:CAA29557.1; PID:G1335034  
 A/Note: the mRNA portion of the sequence corresponds to residues 2092-2157  
 R.Chu, M.L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutzmann, R.; Timpl, R.  
 J. Biol. Chem. 263, 18601-18606, 1988  
 A/Title: Amino acid sequence of the triple-helical domain of human collagen type VI.  
 A/Reference number: A31952; MUID:89066644; PMID:3198591  
 A/Accession: C31952  
 A/Molecule type: mRNA  
 A/Residues: 2038-2373 <CH4>  
 A/Cross-references: GB:J04211; GB:M30778  
 A/Note: parts of this sequence were determined by protein sequencing  
 R.Well, D.; Mattei, M.G.; Passagge, E.; Van Cong, N.; Pribula-Conway, D.; Mann, K.; Deutz  
 Am. J. Hum. Genet. 42, 435-445, 1988  
 A/Title: Cloning and chromosomal localization of human genes encoding the three chains c  
 A/Reference number: A29848; MUID:88161046; PMID:3348212  
 A/Accession: C29848  
 A/Molecule type: mRNA  
 A/Residues: 2092-2151 <WEI>  
 A/Cross-references: GB:M27449; NID:G291919; PIDN:AAA52057.1; PID:G291920  
 A/Note: part of this sequence was determined by protein sequencing  
 R.Jander, R.; Rautenberg, J.; Glanville, R.W.  
 Eur. J. Biochem. 133, 39-46, 1983  
 A/Title: Further characterization of the three polypeptide chains of bovine and human sh  
 A/Reference number: S26510; MUID:83209648; PMID:6852033  
 A/Accession: S26510  
 A/Molecule type: protein  
 A/Residues: 'SAIVAGVAGV' <JAN>  
 A/Note: this sequence cannot be reliably placed and probably represents the results from  
 R.Mayer, U.; Poeschl, E.; Nischt, R.; Specks, U.; Pan, T.C.; Chu, M.L.; Timpl, R.  
 Eur. J. Biochem. 225, 573-580, 1994

A/Title: Recombinant expression and properties of the Kunitz-type protease-inhibitor mod  
 A/Reference number: S48709; MUID:95045506; PMID:7525281  
 A/Accession: S48709  
 A/Molecule type: mRNA  
 A/Residues: 'NRAMIFFLICAGRAAA', 3102-3176 <NAV>  
 A/Note: engineered sequence to allow isolation of the Kunitz-type proteinase inhibitor h  
 R.Arnaud, B.; Merigau, K.; Saladjian, P.; Norris, F.; Norris, K.; Bjorn, S.; Olsen, O.;  
 submitted to the Brookhaven Protein Data Bank, August 1994  
 A/Reference number: A52812; PDB:1KXVT  
 A/Contents: annotation; X-ray crystallography, 1.6 angstroms, residues 3106-3160  
 C/Comment: engineered sequence expressed in *Saccharomyces cerevisiae* strain mt-663  
 C/Comment: Proline and lysines at the third position of the tripeptide repeating unit (s  
 h) are 5-hydroxylated and subsequently O-glycosylated.  
 C/Comment: The fibronectin type III repeat homology domain may be released during proces  
 C/Genetics:  
 A/Genes: GDB:COL6A3  
 A/Cross-references: GDB:119066; OMIM:120250  
 A/Map position: 2q37.3-2q37.3  
 C/Complex: type VI collagen is a heterotrimer of one alpha 1(VI) chain (see PIR:CGHUA),  
 sociations among trimer amino- and carboxyl-terminal domains (with disulfide bonds).  
 C/Function:  
 A/Description: structural component of extracellular tissue microfibrils associated with  
 C/Superfamily: collagen alpha 3(VI) chain; animal kunitz-type proteinase inhibitor homol  
 C/Keywords: alternative splicing; blocked amino end; cell binding; coiled coil; extracel  
 F/1-25/DNA: signal sequence #status predicted <SIG>  
 F/26-3176/Product: collagen alpha 3(VI) chain #status predicted <MAT1>  
 F/26-2037/Product: amino-terminal nonhelical #status predicted <ANH>  
 F/26-30,237-3176/Product: collagen alpha 3(VI) chain, splice form A9/N10(-) #status pred  
 F/327-203/DNA: von Willebrand factor type A repeat homology <W01>  
 F/240-405/DNA: von Willebrand factor type A repeat homology <W02>  
 F/443-608/DNA: von Willebrand factor type A repeat homology <W03>  
 F/637-802/DNA: von Willebrand factor type A repeat homology <W04>  
 F/835-999/DNA: von Willebrand factor type A repeat homology <W05>  
 F/1027-1191/DNA: von Willebrand factor type A repeat homology <W06>  
 F/1123-1394/DNA: von Willebrand factor type A repeat homology <W07>  
 F/1433-1599/DNA: von Willebrand factor type A repeat homology <W08>  
 F/1637-1802/DNA: von Willebrand factor type A repeat homology <W09>  
 F/1836-2005/DNA: von Willebrand factor type A repeat homology <W10>  
 F/2038-2373/DNA: interrupted helical  
 F/2040-2042/Region: cell attachment (R-G-D) motif  
 F/2136-2138/Region: cell attachment (R-G-D) motif  
 F/2148-2150/Region: cell attachment (R-G-D) motif  
 F/2154-2156/Region: cell attachment (R-G-D) motif  
 F/2370-2372/Region: cell attachment (R-G-D) motif  
 F/2374-3176/DNA: carboxyl-terminal nonhelical #status predicted <CNH>  
 F/2400-2571/DNA: von Willebrand factor type A repeat homology <W11>  
 F/2617-2800/DNA: von Willebrand factor type A repeat homology <W12>  
 F/2865-2886/Region: alanine/lysine/proline/threonine/valine-rich repeats  
 F/2987-3072/DNA: fibronectin type III repeat homology <FN3>  
 F/3111-3161/DNA: animal Kunitz-type proteinase inhibitor homology <BPI>  
 F/326/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F/108,116,202,251,2079,2331,2558,2677,2861,3036/Binding site: carbonylate (Asn) (covale  
 F/2087/Disulfide bonds: interchain #status predicted  
 F/2100,2206,2239,2316,2319/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F/2103,2209,2212,2322,2337/Binding site: carboxylate (Lys) (covalent) #status experimen  
 F/2103,2209,2212,2322,2337/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F/3111-3161,3120-3144,3136-3157/Disulfide bonds: #status predicted

Query Match 0.74; Score 8; DB 1; Length 3176;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 IQVGLVQY 209  
 DB 1676 IQVGLVQY 1683

RESULT 64  
 DB1666  
 hypochelated protein TC0766 [imported] - Chlamydia muridarum (strain Nigg)  
 C/Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
 C/Date: 31-Mar-2000 #sequence revision 31-Mar-2000 #ext change 11-May-2000  
 C/Accession: DB1666



R,Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salberg, Nucleic Acids Res. 28, 1397-1406, 2000  
 A>Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39.  
 A:Reference number: AB1500; MUID:20150255; PMID:10684935  
 A:Accession: D81666  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-47 <TET>  
 A:Cross-references: GB:AE002345; GB:AE002160; NID:g7190791; PIDN:AAF39569.1; PID:g719079  
 A:Experimental source: strain N199 (MoPn)  
 A:Genetics: TC0766

Query Match  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALL 1140  
 |||||  
 Db 4 LLLALL 10

RESULT 65  
 C65069  
 hypothetical protein b2858 - *Escherichia coli* (strain K-12)  
 C:Species: *Escherichia coli*  
 C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C:Accession: C65069  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC  
 A.; Rose, D.J.; Mau, B.; Shaoh, Y.  
 Science 277, 1453-1462, 1997  
 A>Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: C65069  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-73 <BLAT>  
 A:Cross-references: GB:AE000369; GB:U00096; NID:g2367168; PIDN:AACT5897.1; PID:g1789222;  
 A:Experimental source: strain K-12, substrain MG1655

Query Match  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1158 EEKREK 1164  
 |||||  
 Db 32 EEKREK 38

RESULT 66  
 T10456  
 dermaseptin B3 precursor - two-colored leaf frog  
 C:Species: *Phyllomedusa bicolor* (two-colored leaf frog)  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
 C:Accession: T10456  
 R:Charpentier, S.; Amiche, M.; Meister, J.; Vouille, V.; Le Caer, J.P.; Niccolas, P.; Del  
 J. Biol. Chem. 273, 14690-14697, 1998  
 A>Title: Structure, synthesis, and molecular cloning of dermaseptins B, a family of skin  
 A:Reference number: Z17027; MUID:98287974; PMID:9614066  
 A:Accession: T10456  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-74 <CHA>  
 A:Cross-references: EMBL:Y16564; NID:g3256036; PIDN:CAA76288.1; PID:g3256037  
 C:Superfamily: dermaseptin precursor; dermorphin precursor amino-terminal homology  
 F.1-22/Domains: signal sequence #status predicted <SIG>  
 F.23-74/Product: dermaseptin B3 #status predicted <MAT>

Query Match  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1157 EEKREK 1163  
 |||||  
 Db 23 EEKREK 29

RESULT 67  
 I51527  
 integrin alpha 5 subunit - African clawed frog (fragment)  
 C:Species: *Xenopus laevis* (African clawed frog)  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 29-Sep-1999  
 C:Accession: I51527  
 R:Whitaker, C.A.; Desimone, D.W.  
 Development 117, 1239-1249, 1993  
 A>Title: Integrin alpha subunit mRNAs are differentially expressed in early *Xenopus* embr  
 A:Reference number: I51524; MUID:94008528; PMID:8404528  
 A:Accession: I51527  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-76 <WHI>  
 A:Cross-references: GB:U10191; NID:g214544; PIDN:AAA16249.1; PID:g214545  
 C:Superfamily: integrin alpha-2b chain

Query Match  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 GRVYVYL 526  
 |||||  
 Db 36 GRVYVYL 42

RESULT 68  
 I51529  
 integrin alpha V-like subunit - African clawed frog (fragment)  
 C:Species: *Xenopus laevis* (African clawed frog)  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 29-Sep-1999  
 C:Accession: I51529  
 R:Whitaker, C.A.; Desimone, D.W.  
 Development 117, 1239-1249, 1993  
 A>Title: Integrin alpha subunit mRNAs are differentially expressed in early *Xenopus* embr  
 A:Reference number: I51524; MUID:94008528; PMID:8404528  
 A:Accession: I51529  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-77 <WHI>  
 A:Cross-references: GB:U10190; NID:g214548; PIDN:AAA16251.1; PID:g214549  
 C:Superfamily: integrin alpha-2b chain

Query Match  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 GRVYVYL 526  
 |||||  
 Db 36 GRVYVYL 42

RESULT 69  
 D70640  
 hypothetical protein RV0689c - *Mycobacterium tuberculosis* (strain H37RV)  
 C:Species: *Mycobacterium tuberculosis*  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: D70640  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.  
 A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: D70640  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA



A:Residues: 1-84 <COL>  
A:Cross-references: GB:Z84395; GB:A1123456; NID:g91261698; PIDN:CAB06452.1; PID:e293110;  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV0689c

Query Match 0.6%; Score 7; DB 2; Length 84;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1049 RCHLGOL 1055  
DB 78 RCHLGOL 84

RESULT 70  
G82776  
ACP XF0672 (imported) - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 01-Dec-2000  
C:Accession: G82776  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A>Note: For a complete list of authors see reference number A59328 below  
A:Accession: G82776  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-85 <STM>  
A:Cross-references: GB:AE003911; GB:AE003849; NID:g9105548; PIDN:AAF33482.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kempfer, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, H  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmetti, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Saneili, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF0672  
C:Superfamily: acyl carrier protein; acyl carrier protein homology  
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F:8-79/Domain: acyl carrier protein homology <ACP>  
F:43/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 0.6%; Score 7; DB 2; Length 85;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 414 MALDEDF 420  
DB 51 MALDEDF 57

RESULT 71  
B91093  
hypothetical protein Ecs3714 (imported) - Escherichia coli (strain O157:H7, substrain RI  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: B91093  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gesawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: B91093  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-86 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA87137.1; PID:g13363186; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain R1MD 050952  
C:Genetics:  
A:Gene: Ecs3714

Query Match 0.6%; Score 7; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1158 EKKREK 1164  
DB 49 EKKREK 55

RESULT 72  
G85938  
hypothetical protein Z4179 (imported) - Escherichia coli (strain O157:H7, substrain EDL5  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G85938  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaeser, J.D.; Rose, D.J.; Mayhev  
iller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85938  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-86 <STO>  
A:Cross-references: GB:AE005174; NID:g12517354; PIDN:AA657971.1; GSPDB:GN00145; UMGF:Z41  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z4179

Query Match 0.6%; Score 7; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1158 EKKREK 1164  
DB 49 EKKREK 55

RESULT 73  
I49515  
B144 protein B - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: I49515  
R:Tsuge, I.; Shen, F.  
Immunogenetics 26, 378-380, 1987  
A:Title: A gene in the H-2S: H-2D interval of the major histocompatibility complex which  
A:Reference number: I49514; MUID:88031493; PMID:3117682  
A:Accession: I49515  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-89 <RES>  
A:Cross-references: GB:M18187; NID:g192097; PIDN:AAA37273.1; PID:g192099

Query Match 0.6%; Score 7; DB 2; Length 89;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLL 1137  
DB 36 LGGLLLL 42

RESULT 74

T03285  
 anther-specific protein - rice  
 C:Species: Oryza sativa (rice)  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #ext\_change 24-Mar-1999  
 C:Accession: T03285  
 R:Lee, J.Y.K.; Hodges, T.K.  
 A:Submitted to the EMBL Data Library, July 1994  
 A:Description: Genomic DNA sequence of a rice anther-specific gene.  
 A:Reference number: Z14882  
 A:Accession: T03285  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-94 <LEE>  
 A:Cross-references: EMBL:U12171; NID:g607894; PID:g607895  
 A:Experimental source: strain IR54  
 C:Genetics:  
 A:Gene: R1S2

Query Match 0.6%; Score 7; DB 2; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 468 DGAVRVA 474  
 |||||  
 DB 31 DGAVRVA 37

## RESULT 75

A64146  
 YajC protein homolog HI0241 - Haemophilus influenzae (strain Rd KW20)  
 C:Species: Haemophilus influenzae  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #ext\_change 28-Jul-2000  
 C:Accession: A64146  
 R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geognagen, N.S.M.  
 Science 269, 496-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630; PMID:7542800  
 A:Accession: A64146  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-97 <RIGR>  
 A:Cross-references: GB:U32710; GB:LA2023; NID:g1573200; PIDN:AA21909.1; PID:g1573206; T  
 C:Superfamily: YajC protein  
 C:Keywords: transmembrane protein  
 F:8-24/Domain: transmembrane #status predicted <TMM>

Query Match 0.6%; Score 7; DB 1; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 89;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1055 LAKGTEV 1061  
 |||||  
 DB 42 LAKGTEV 48

Search completed: July 16, 2003, 08:34:15  
 Job time : 34 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:28:30 ; Search time 52 Seconds

(without alignments)  
4624.173 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167  
Sequence: 1 MELPFVTHLPVFLTGLC.....GFAHKKIPBEKREKLEQ 1167

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	1.1	1171	13	042094
2	11	0.9	191	6	029124
3	11	0.9	383	2	09R786
4	11	0.9	823	4	08WY18
5	10	0.9	780	13	006271
6	9	0.8	66	10	0943V3
7	9	0.8	84	16	0932R3
8	9	0.8	236	9	064016
9	9	0.8	236	16	032005
10	9	0.8	244	16	09RJ10
11	9	0.8	399	16	08XSH5
12	9	0.8	757	10	094LU9
13	9	0.8	1086	4	096H81
14	9	0.8	1160	11	09SR200
15	9	0.8	1161	11	09WTV4
16	9	0.8	1187	13	098TF0

17	9	0.8	1196	13	098TF1	098TF1 cyprinus ca
18	8	0.7	59	4	09UR7	09UR7 homo sapien
19	8	0.7	77	9	09MC41	09MC41 bacterioph
20	8	0.7	97	4	000453	000453 homo sapien
21	8	0.7	103	16	08XEP2	08XEP2 salmonella
22	8	0.7	105	2	09EY13	09EY13 klebsiella
23	8	0.7	112	11	08RIE9	08RIE9 mus musculu
24	8	0.7	119	11	09OX14	09OX14 rattus norv
25	8	0.7	127	16	09RSB5	09RSB5 streptomyce
26	8	0.7	136	16	09KN21	09KN21 vibrio chol
27	8	0.7	141	11	061692	061692 mus musculu
28	8	0.7	146	13	08G930	08G930 bothrops in
29	8	0.7	169	16	08ZLM7	08ZLM7 salmonella
30	8	0.7	169	16	08Z1W9	08Z1W9 salmonella
31	8	0.7	181	13	090Y12	090Y12 crocalus du
32	8	0.7	181	13	090Y11	090Y11 crocalus du
33	8	0.7	185	16	097LC8	097LC8 clostridium
34	8	0.7	240	17	09IBV7	09IBV7 aeropyrum p
35	8	0.7	258	16	09RK95	09RK95 streptomyce
36	8	0.7	265	13	09PM56	09PM56 bothrops ja
37	8	0.7	265	13	08OC91	08OC91 bothrops in
38	8	0.7	269	5	09UI68	09UI68 leishmania
39	8	0.7	281	2	053905	053905 streptomyce
40	8	0.7	281	16	093122	093122 streptomyce
41	8	0.7	287	2	093QX2	093QX2 corynebacte
42	8	0.7	294	2	09XCW9	09XCW9 rhodobacter
43	8	0.7	298	16	084105	084105 chlamydia t
44	8	0.7	299	2	09R9K7	09R9K7 paracoccus
45	8	0.7	299	10	094K03	094K03 arabidopsis
46	8	0.7	309	2	005656	005656 bacillus ce
47	8	0.7	348	4	08TES5	08TES5 homo sapien
48	8	0.7	360	10	09M146	09M146 arabidopsis
49	8	0.7	383	5	016339	016339 caenorhabdi
50	8	0.7	385	5	09NFP1	09NFP1 leishmania
51	8	0.7	405	10	024416	024416 fragaria an
52	8	0.7	405	10	094FT6	094FT6 fragaria an
53	8	0.7	415	16	091573	091573 pseudomonas
54	8	0.7	433	16	08YOR5	08YOR5 anabaena sp
55	8	0.7	444	10	004611	004611 arabidopsis
56	8	0.7	471	16	08XZT4	08XZT4 ralsionia s
57	8	0.7	491	2	09REI8	09REI8 rhodobacter
58	8	0.7	491	2	09K559	09K559 rhodobacter
59	8	0.7	491	11	09WUD0	09WUD0 mus musculu
60	8	0.7	491	11	064584	064584 rattus norv
61	8	0.7	497	13	073804	073804 fuqu rubrip
62	8	0.7	518	10	09A541	09A541 cryza sativ
63	8	0.7	573	4	08WY13	08WY13 homo sapien
64	8	0.7	577	16	08R6N2	08R6N2 thermocae
65	8	0.7	596	17	08TRM8	08TRM8 methanosarc
66	8	0.7	607	10	039775	039775 gnetum gnet
67	8	0.7	614	16	031615	031615 bacillus su
68	8	0.7	622	13	091570	091570 xenopus lae
69	8	0.7	636	5	017795	017795 caenorhabdi
70	8	0.7	636	16	09CWV6	09CWV6 pasteurella
71	8	0.7	680	11	070491	070491 mus musculu
72	8	0.7	685	16	08Z9B5	08Z9B5 salmonella
73	8	0.7	704	4	096PV1	096PV1 homo sapien
74	8	0.7	739	12	09INI9	09INI9 kadipito vi
75	8	0.7	767	10	09CAD3	09CAD3 arabidopsis
76	8	0.7	849	16	0911A8	0911A8 pseudomonas
77	8	0.7	952	10	08S920	08S920 cryza sativ
78	8	0.7	1036	6	09TU44	09TU44 canis faml
79	8	0.7	1036	6	09TUN8	09TUN8 canis faml
80	8	0.7	1036	6	09TUN6	09TUN6 sus scrofa
81	8	0.7	1049	5	08SY51	08SY51 dtrosophila
82	8	0.7	1161	11	09QY57	09QY57 rattus norv
83	8	0.7	1500	10	09Z084	09Z084 arabidopsis
84	8	0.7	2109	4	012789	012789 homo sapien
85	8	0.7	2221	5	09UIM1	09UIM1 dtrosophila
86	8	0.7	2222	5	097394	097394 dtrosophila
87	8	0.7	4899	5	09VR91	09VR91 dtrosophila
88	7	0.6	22	6	09TRC4	09TRC4 canis faml
89	7	0.6	25	4	09UL44	09UL44 homo sapien

90 7 0.6 47 16 Q9PJ2 chlamydia m  
91 7 0.6 56 11 O09128 mus musculus  
92 7 0.6 73 16 P76640 escherichia  
93 7 0.6 77 13 O06276 xenopus lae  
94 7 0.6 78 6 Q9N241 o9n241 saintiri bol  
95 7 0.6 79 5 Q8S762 Q8S762 drosophila  
96 7 0.6 79 16 Q8Y0J1 Q8Y0J1 ralsstonia s  
97 7 0.6 81 11 O08844 s O08844 mus musculus  
98 7 0.6 84 15 O91182 O91182 human immun  
99 7 0.6 84 16 P95035 P95035 mycobacteri  
100 7 0.6 85 16 Q9PFI5 Q9PFI5 xyella fas

## ALIGNMENTS

RESULT 1  
042094 PRELIMINARY; PRT; 1171 AA.  
ID 042094  
AC 042094  
DT 01-JAN-1998 (TRENBLREL. 05, Created)  
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)  
DE 01-MAR-2002 (TRENBLREL. 20, Last annotation update)  
OS ALPHA1 integrin.  
OC Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OC NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=GIZZARD;  
RA MEDLINE=97476270; PubMed=9334246;  
RA Obara H., Hayashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,  
RA Sobue K.;  
RT "Smooth muscle cell phenotype-dependent transcriptional regulation of  
RT the alpha1 integrin gene.";  
RL J. Biol. Chem. 272:26643-26651(1997).  
DR EMBL; AB000470; BAA23160.1; -;  
DR EMBL; AB000471; BAA23161.1; -;  
DR HSSP; P17301; IAOX.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01839; FG-GAP; 5.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 4.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS50234; VWF\_A; 1.  
DR Integrin.  
DR Integrin.  
SQ SEQUENCE 1171 AA; 130228 MW; B5054MD65F09736F CRC64;  
Query Match 1.1%; Score 13; DB 13; Length 1171;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSNITPW 181  
DB 164 VIVLDGNSNITPW 176

RESULT 2  
029124 PRELIMINARY; PRT; 191 AA.  
ID 029124  
AC 029124  
DT 01-NOV-1996 (TRENBLREL. 01, Created)  
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)  
DE VLA-2 (Fragment).  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OC NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=ENDOTHELIAL CELLS;  
RX MEDLINE=95036279; PubMed=7949129;  
RA Bahou W.F., Potter C.L., Mirza H.;  
RT "The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific  
RT recognition sequence for endothelial cell attachment and spreading;  
RT molecular and functional characterization.";  
RL Blood 84:3734-3741(1994).  
DR EMBL; Z12137; CAA78125.1; -;  
DR HSSP; P17301; IAOX.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS50234; VWF\_A; 1.  
FT NON\_TER 1 1  
FT NON\_TER 191 191  
SQ SEQUENCE 191 AA; 20896 MW; 2E228B472EC699D8 CRC64;  
Query Match 0.9%; Score 11; DB 6; Length 191;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282  
DB 92 VVVTGESHHDG 102

RESULT 3  
09R7B6 PRELIMINARY; PRT; 383 AA.  
ID 09R7B6  
AC 09R7B6  
DT 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)  
DE Carbon starvation protein (Fragment).  
GN CSG.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OC NCBI\_TaxID=210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CCUG 17874;  
RA Beddek A.J., O'Toole P.W.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U81796; AAD00522.1; -;  
DR InterPro; IPR003706; CstA.  
DR Pfam; PF02554; CstA; 1.  
FT NON\_TER 383 383  
SQ SEQUENCE 383 AA; 41074 MW; C8F75BCBA1ADE9F CRC64;  
Query Match 0.9%; Score 11; DB 2; Length 383;  
Best Local Similarity 100.0%; Pred. No. 0.046;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 LMIILGSLVGG 1133  
DB 118 LMIILGSLVGG 128

RESULT 4  
08WY18 PRELIMINARY; PRT; 823 AA.  
ID 08WY18  
AC 08WY18  
DT 01-MAR-2002 (TRENBLREL. 20, Created)  
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)  
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
DE MSTP018.  
GN MST018.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=AOA;  
 RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,  
 Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,  
 Sun Y.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,  
 Gao R.L., Qiang B.O., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;  
 RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF111799; AAL39001.1; -  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR Pfam: PF01839; FG-GAP; 4.  
 DR PRINTS: PR01185; INTEGRIN.  
 DR SMART: SM00191; Int\_alpha; 4.  
 SQ SEQUENCE 823 AA; 92672 MW; DEAE78079DCD4925 CRC64;

Query Match 0.9%; Score 11; DB 4; Length 823;  
 Best Local Similarity 100.0%; Pred. No. 0.09;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LCGLLALLLV 1141  
 DB 785 LCGLLALLLV 795

## RESULT 5

AC Q06271; PRELIMINARY; PRT; 780 AA.  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE Integrin alpha-2 subunit (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Meng F., Desimone D.W.;  
 RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 95-168 FROM N.A.  
 RX MEDLINE=94008528; PubMed=8404528;  
 RA Whitaker C.A., Desimone D.W.;  
 RT "Integrin alpha subunit mRNAs are differentially expressed in early  
 RT Xenopus embryos.";  
 RL Development 117:1239-1249 (1993).  
 DR EMBL: L43058; AAA69770.1; -  
 DR EMBL: L10186; AAA16246.1; -  
 DR HSP; P11215; IABX.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR Pfam: PF01839; FG-GAP; 4.  
 DR PRINTS: PR01185; INTEGRIN.  
 DR SMART: SM00191; Int\_alpha; 4.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 FT NON TER 1  
 SQ SEQUENCE 780 AA; 87017 MW; 9518B18C2B6B637 CRC64;

Query Match 0.9%; Score 10; DB 13; Length 780;  
 Best Local Similarity 100.0%; Pred. No. 0.92;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 GEOIGSYFGS 488  
 DB 86 GEOIGSYFGS 95

RESULT 6  
 ID Q943Y3 PRELIMINARY; PRT; 66 AA.  
 AC Q943Y3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE P0552C05.18 protein.  
 GN P0552C05.18.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPONBARE;  
 RA Sasaki T., Matsunoto T., Yamamoto K.;  
 RT "Oryza sativa niphonbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0552C05.";  
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AP002873; BAB64033.1; -  
 SQ SEQUENCE 66 AA; 7136 MW; 196108BABC7972E2 CRC64;

Query Match 0.8%; Score 9; DB 10; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 866 ECAPSAHA 874  
 DB 11 ECAPSAHA 19

RESULT 7  
 ID Q922R3 PRELIMINARY; PRT; 84 AA.  
 AC Q922R3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein RA0386.  
 GN RA0386 OR SMA0726.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Rhizobium psyma (megaplasmid 1).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 Barloy-Hudler F., Bowser L., Capela D., Galibert F., Gouzy J.,  
 Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,  
 Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Sinorhizobium meliloti pSyma megaplasmid.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).  
 DR EMBL: AE007230; AAK65044.1; -  
 KW Hypothetical protein; plasmid; Complete proteome.  
 SQ SEQUENCE 84 AA; 9792 MW; 5A847AB033610F9F CRC64;

Query Match 0.8%; Score 9; DB 16; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 RLSPRLRL 729  
 DB 58 RLSPRLRL 66

## RESULT 8

```

064016          PRELIMINARY;          PRT;          236 AA.
ID 064016;
AC 064016;
DT 01-ANG-1998 (TrEMBLrel. 07, Created)
DT 01-ANG-1998 (TrEMBLrel. 07, Last sequence update)
DE Putative lipoprotein.
GN YOKB.
OS Bacteriophage SPBc2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxId=66797;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.,
RA Karimata D.;
RT "The complete nucleotide sequence of the Bacillus subtilis SPbetac2
RT prophage.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020713; AAC12975.1; -.
KW Lipoprotein.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26C0CF4F4AC CRC64;

Query Match          0.8%; Score 9; DB 9; Length 236;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 PEEKREK 1164
Db 209 PEEKREK 217

RESULT 9
032005          PRELIMINARY;          PRT;          236 AA.
ID 032005;
AC 032005;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE YOKB protein.
GN YOKB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Betero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Bourcier L., Brans A., Braun M., Bridgell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.U., Comercon I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Duesterhoft A., Enrich S.D., Emmerson P.T.,
RA Eutian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
RA Joris B., Kazanata D., Kasanara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Kirogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber B., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleier S., Schoefer R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

```

```

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipac A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99115; CAB14083.1; -.
KW Complete proteome.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26C0CF4F4AC CRC64;

Query Match          0.8%; Score 9; DB 16; Length 236;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 PEEKREK 1164
Db 209 PEEKREK 217

RESULT 10
09RJL0          PRELIMINARY;          PRT;          244 AA.
ID 09RJL0;
AC 09RJL0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Putative secreted protein.
GN SC00378 OR SCF62.04C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Ruter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147 (2002).
DR EMBL; AL121855; CAB58319.1; -.
SQ SEQUENCE 244 AA; 25456 MW; 7C5575F6710986B CRC64;

Query Match          0.8%; Score 9; DB 16; Length 244;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 ADVAVGAPL 573
Db 171 ADVAVGAPL 179

RESULT 11
08XSH5          PRELIMINARY;          PRT;          399 AA.
ID 08XSH5;
AC 08XSH5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transport transmembrane protein.
GN RSP0499 OR RS00378.

```

OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681679; PubMed=11823852;  
 RA Salanoubat M., Genin S., Arciguenave F., Gouzy J., Mangenot S.,  
 RA Arlet M., Billault A., Brotier P., Camus J.C., Cartolico L.,  
 RA Chandler C., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Weisenbach J., Boucher C.A.,  
 RA "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RT Nature 415:497-502 (2002).  
 RL EMBL: AL646079; CAD17650.1; .  
 DR InterPro: IPR001179; FKBP\_PPIase.  
 DR PROSITE: PS00453; FKBP\_PPIase\_1; UNKNOWN\_1.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 399 AA; 42148 MW; 096BB97E83DCA7D1 CRC64;

Query Match 0.8%; Score 9; DB 16; Length 399;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLTLALLV 1141  
 DB 161 GLLTLALLV 169

RESULT 12

O94LU9 PRELIMINARY; PRT; 757 AA.  
 ID O94LU9;  
 AC O94LU9;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Putative  
 DE UDP-N-acetylglucosaminyl-D-glutamate-2,6-diaminopimelate  
 DE 1lgase.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhacoidae; Oryzae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,  
 RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tsirlin T.,  
 RA Riggs P., Haiso J., Zismann V., Blunt S., Pai G., Vanhaken S.E.,  
 RA Utterback T.R., Feldblum T.V., Quackenbush J., Galzberg S.L.,  
 RA White O., Frazer C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSJNBa0082M15 genomic sequence";  
 RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC020666; AKA4503.1; .  
 DR InterPro: IPR000713; Mur\_1lgase.  
 DR InterPro: IPR004101; Mur\_1lgase\_C.  
 DR Pfam: PF01225; Mur\_1lgase; 1.  
 DR Pfam: PF02875; Mur\_1lgase\_C; 1.  
 SQ SEQUENCE 757 AA; 82194 MW; 8EF030AFAD51F80 CRC64;

Query Match 0.8%; Score 9; DB 10; Length 757;  
 Best Local Similarity 100.0%; Pred. No. 9.7;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 567 VAVGAPLED 575  
 DB 543 VAVGAPLED 551

RESULT 13  
 ID Q96HB1 PRELIMINARY; PRT; 1086 AA.  
 AC Q96HB1;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Unknown (protein for MGC:1714).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph;  
 RA Strussberg R.;  
 RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: BC008777; AAH08777.1; .  
 DR InterPro: IPR00413; Integrin\_alpha.  
 DR InterPro: IPR02035; VWF\_A.  
 DR Pfam: PF01839; FG-GAP; 5.  
 DR Pfam: PF00357; Integrin\_A; 1.  
 DR Pfam: PF00092; vwa; 1.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; UNKNOWN\_1.  
 DR PROSITE: PS50234; VWF\_A; 1.  
 SQ SEQUENCE 1086 AA; 119223 MW; F6FP2546E8C632F9 CRC64;

Query Match 0.8%; Score 9; DB 4; Length 1086;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 566 DVAVGAPLE 574  
 DB 455 DVAVGAPLE 463

RESULT 14

O9R200 PRELIMINARY; PRT; 1160 AA.  
 ID O9R200;  
 AC O9R200;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Integrin alpha L.  
 OS ITGAL.  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=SPLEEN;  
 RA R.Z., Teuscher C.;  
 RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF065901; AAD25884.1; .  
 DR HSSP: P20701; ILPA.  
 DR MGD: MGI:96606; Itgal.  
 DR InterPro: IPR00413; Integrin\_alpha.  
 DR InterPro: IPR02035; VWF\_A.  
 DR Pfam: PF01839; FG-GAP; 5.  
 DR Pfam: PF00357; Integrin\_A; 1.  
 DR Pfam: PF00092; vwa; 1.  
 DR PRINTS: PR01185; INTEGRINA.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 DR SMART: SM00191; Itc\_alpha; 5.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE: PS50234; VWF\_A; 1.  
 SQ SEQUENCE 1160 AA; 128127 MW; A33C531B139F1PAD CRC64;

Query Match 0.8%; Score 9; DB 11; Length 1160;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574  
 |||||  
 DB 536 DVAVGAPLE 544

## RESULT 15

Q9WTV4 PRELIMINARY; PRT; 1161 AA.

DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Integrin alpha L.  
 GN ITGAL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBA/2J; TISSUE=SPLEN;  
 RA Ma R.Z.; Teuscher C.;  
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
 DR HSSP; P20701; ILFA.  
 DR MGD; MG1:96606; Itgal.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00327; vwa; 1.  
 DR SMART; SM00327; vwa; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 SQ SEQUENCE 1161 AA; 128240 MW; 86B102F7B209E431 CRC64;

Query Match 0.8%; Score 9; DB 11; Length 1161;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574  
 |||||  
 DB 536 DVAVGAPLE 544

## RESULT 16

Q98TF0 PRELIMINARY; PRT; 1187 AA.

AC 098TF0; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE CD11-2.  
 GN CIA2.

OS Cyprinus carpio (Common carp).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Cyprinus.  
 OX NCBI\_TaxID=7962;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PERITONEAL EXUDATE CELLS;  
 RA Kimura M., Fujiki K., Nakao M.;  
 RT Molecular cloning of a leukocyte integrin from the common carp.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB048537; BAB39135.1; -.  
 DR HSSP; P20701; ILFA.  
 DR InterPro; IPR001969; Asprotease\_site.  
 DR InterPro; IPR001969; Asprotease\_site.  
 DR InterPro; IPR000413; Integrin\_alpha.

DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00327; vwa; 1.  
 DR SMART; SM00327; vwa; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 SQ SEQUENCE 1187 AA; 131778 MW; 85EDC7CA8B6B1C59 CRC64;

Query Match 0.8%; Score 9; DB 13; Length 1187;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 GEOIGSYFG 487  
 |||||  
 DB 454 GEOIGSYFG 462

## RESULT 17

Q98TF1 PRELIMINARY; PRT; 1196 AA.

AC 098TF1; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE CD11-1.  
 GN CIA1.

OS Cyprinus carpio (Common carp).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Cyprinus.  
 OX NCBI\_TaxID=7962;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PERITONEAL EXUDATE CELLS;  
 RA Kimura M., Nakao M., Mura C., Fujiki K., Yano T.;  
 RT Molecular cloning of a leukocyte integrin from the common carp.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB048536; BAB39134.1; -.  
 DR HSSP; P20701; ILFA.  
 DR InterPro; IPR001969; Asprotease\_site.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PRO1185; INTEGRINA.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00327; vwa; 1.  
 DR SMART; SM00327; vwa; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 SQ SEQUENCE 1196 AA; 132477 MW; 9369C807E7DCA53B CRC64;

Query Match 0.8%; Score 9; DB 13; Length 1196;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 GEOIGSYFG 487  
 |||||  
 DB 454 GEOIGSYFG 462

## RESULT 18

Q9UR7 PRELIMINARY; PRT; 59 AA.

AC 09UR7; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)



```

DE LST-1/K protein.
GN LST-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RA Rollinger-Holinger I., Eibl B., Pauly M., Weiss E.;
RT "LST1 gene with complex alternative splicing and immunosuppressive
function." (DEC-1998) to the EMBL/GenBank/DBJ databases.
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y18487; CAB59905.1; -.
SQ SEQUENCE 59 AA; 6122 MW; FC2FF156D576C486 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1131 LGGLLLA 1138
Db 16 LGGLLLA 23

RESULT 19
O9MC41 PRELIMINARY; PRT; 77 AA.
ID O9MC41.
AC O9MC41.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Oref1.
GN Oref1.
OS Bacteriophage D3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=31535;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20042341; PubMed=10572124;
RA Gilakjan Z.A., Kropinski A.M.;
RT "Cloning and analysis of the capsid morphogenesis genes of Pseudomonas
aeruginosa bacteriophage D3; another example of protein chain malty";
RL J. Bacteriol. 181:7221-7227(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20485557; PubMed=11029426;
RA Kropinski A.M.;
RT "Sequence of the Genome of the Temperate, Serotype-Converting,
Pseudomonas aeruginosa Bacteriophage D3.";
RL J. Bacteriol. 183:6066-6074(2000).
DR EMBL; AF165214; AAF80767.1; -.
SQ SEQUENCE 77 AA; 8266 MW; 39DAB91077F7234B CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 9; Length 77;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 756 TVTFALDN 763
Db 42 TVTFALDN 49

RESULT 20
O00453 PRELIMINARY; PRT; 97 AA.
ID O00453.
AC O00453.
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LST1 protein.
GN LST1.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98035883; PubMed=9367684;
RA de Baey A., Fellerhoff B., Mater S., Martinozzi S., Weidie U.,
RA Weiss E.H.;
RT "Complex expression pattern of the TNF region gene LST1 through
differential regulation, initiation, and alternative splicing.";
RL Genomics 45:581-600(1997).
DR EMBL; AF000425; AAB86999.1; -.
SQ SEQUENCE 97 AA; 10822 MW; AA03C761E787AF94 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 97;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1131 LGGLLLA 1138
Db 16 LGGLLLA 23

RESULT 21
O8XEP2 PRELIMINARY; PRT; 103 AA.
ID O8XEP2.
AC O8XEP2.
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative septum formation initiator (Hypothetical protein
SY3056).
GN YGBO OR STM2931 OR STY3056.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneila.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Malyaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Doud L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AE008833; AAL21811.1; -.
DR EMBL; AL627276; CAD06037.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11575 MW; F3B55AB77E8AC6C CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 103;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



#

## RESULT 26

O9KNZ1

PRELIMINARY; PRT; 136 AA.

AC O9KNZ1; 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Ribosomal protein L16.  
 GN VCS589.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 NC NCB1\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N1691 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952201;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unsay L.A.,  
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 Ermolaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,  
 McDonald L., Uitterlinden T., Fleischmann R.D., Mierman W.C., White O.,  
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 Fraser C.M.;  
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae."  
 RT Nature 406:477-483 (2000).  
 RL EMBL; AE004326; AAF95730.1; -.  
 DR TIGR; VC2589; -.  
 DR InterPro; IPR000114; Ribosomal\_L16.  
 DR Pfam; PF00252; Ribosomal\_L16; 1.  
 DR PRINTS; PR00060; Ribosomal\_L16.  
 DR TIGRFAMs; TIGR01164; rplp\_bact; 1.  
 DR PROSITE; PS00586; RIBOSOMAL\_L16\_1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 136 AA; 15551 MW; 3F74B546B61F917D CRC64;

## Query Match

Best Local Similarity 0.7%; Score 8; DB 16; Length 136;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1055 LAKGTEVS 1062  
 DB 20 LAKGTEVS 27

## RESULT 27

O61692

PRELIMINARY; PRT; 141 AA.

AC O61692; 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE Putative heat stable antigen.  
 GN HSA-C.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBA X C57BL/6; TISSUE=SPLEEN;  
 RX MEDLINE=91209380; PubMed=2019286;  
 RA Wenger R.H., Ayane M., Bose R., Koehler G., Nielsen P.J.;  
 RT "The genes for a mouse hematopoietic differential marker called the  
 heat-stable antigen."  
 RT Eur. J. Immunol. 21:1039-1046 (1991).  
 DR EMBL; X56486; CAA39843.1; -.  
 SQ SEQUENCE 141 AA; 15515 MW; E4BFB428ADC03C69 CRC64;

Query Match 0.7%; Score 8; DB 11; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLLIALL 1140  
 DB 12 GLLLIALL 19

## RESULT 28

O8G930

PRELIMINARY; PRT; 146 AA.

AC O8G930; 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Bradykinin-potentiating protein.  
 OS Bothrops insularis (Island jararaca) (Queimada jararaca).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Lepidodonta; Squamata; Chordata; Serpentes; Colubroidae;  
 OC Viperae; Crotalinae; Bothrops.  
 NC NCB1\_TaxID=8723;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=VENOM GLAND;  
 RA Junqueira-de-Azevedo I.L.M., Ho P.L.;  
 RT "A survey of gene expression and diversity in the venom glands of the  
 pitviper Bothrops insularis through the generation of Expressed  
 Sequence Tags (ESTs)."  
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF490532; AAM09691.1; -.  
 SQ SEQUENCE 146 AA; 16088 MW; F58244C4F24B1799 CRC64;

## Query Match

Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLLIALL 1140  
 DB 10 GLLLIALL 17

## RESULT 29

O8ZLM7

PRELIMINARY; PRT; 169 AA.

AC O8ZLM7; 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Peptide deformylase (EC 3.5.1.31).  
 GN DEF OR STM3406.  
 OS Salmonella typhimurium.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 NC NCB1\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2."  
 RT Nature 413:852-856 (2001).  
 RL EMBL; AE008857; AAL2269.1; -.  
 DR InterPro; IPR000181; pep\_deformylase.  
 DR Pfam; PF01327; pep\_deformylase; 1.  
 DR PRINTS; PR01576; PDBFORMYLASE.  
 DR Prodom; PD003844; pep\_deformylase; 1.  
 DR TIGRFAMs; TIGR00079; pepc\_deformyl; 1.  
 KW Hydrolase; Complete proteome.  
 SQ SEQUENCE 169 AA; 19282 MW; 3181AB2F6BF02765 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 8; DB 16; Length 169;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 LVGKLFID 198  
DB 138 LVGKLFID 145

## RESULT 30

Q821W9 PRELIMINARY; PRT; 169 AA.  
AC Q821W9;  
DT 01-MAR-2002 (TRENBLREL. 20, Created)  
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)  
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
DE Polypeptide deformylase (EC 3.5.1.31).  
GN FMS OR STY4391.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; Pubmed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Fellwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gea P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
RT "Complete genome sequence of a multiple drug resistant *Salmonella* enterica serovar Typhi CT18.";  
RL EMBL: AL627282; CAD09179.1; -;  
DR InterPro: IPR000181; Pep.deformylase.  
DR Pfam: PF01327; Pep.deformylase; 1.  
DR PRINTS: PR01576; PDEFORMLAS.  
DR ProDom: PD003844; Pep.deformylase; 1.  
DR TIGRFAMs: TIGR00079; pepc.deformyl; 1.  
KW Hydroxylase; Complete proteome.

SO SEQUENCE 169 AA; 19281 MW; 618BA12F6BF02ADE CRC64;

Query Match  
Best Local Similarity 100.0%; Score 8; DB 16; Length 169;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 LVGKLFID 198  
DB 138 LVGKLFID 145

## RESULT 31

Q90Y12 PRELIMINARY; PRT; 181 AA.  
AC Q90Y12;  
DT 01-DEC-2001 (TRENBLREL. 19, Created)  
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)  
DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)  
DE Bradykinin potentiating peptide and C-type natriuretic peptide isoform 1.  
GN Crocatus durissus terrificus (South American rattlesnake).  
OS Eukaryota; Metazoa; Chordata; Cranialata; Euteleostomi;  
OC Lepidodermata; Squamata; Scleroglossa; Serpentes; Colubroidae;  
OC Viperidae; Crocaltinae; Crocatus.  
OX NCBI\_TaxID=8732;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=VENOM GLAND;

RA Hayashi M.A.F., Radis-Baptista G., Barbosa S.R., Yamane T., Camargo A.C.M.;  
RT "Crocatus durissus terrificus bradykinin potentiating peptide precursor.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF308593; AAL09426.1; -;  
DR InterPro: IPR000663; Natr\_peptide.  
DR Pfam: PF00212; ANP; 1.  
DR PROSITE: PS00263; NATRIURETIC PEPTIDE; UNKNOWN 1.  
SO SEQUENCE 181 AA; 18560 MW; 7B5ADC5B9372D07F CRC64;

Query Match  
Best Local Similarity 100.0%; Score 8; DB 13; Length 181;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLTALL 1140  
DB 10 GLLTALL 17

## RESULT 32

Q90Y11 PRELIMINARY; PRT; 181 AA.  
AC Q90Y11;  
DT 01-DEC-2001 (TRENBLREL. 19, Created)  
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)  
DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)  
DE Bradykinin potentiating peptide and C-type natriuretic peptide isoform 2.  
GN Crocatus durissus terrificus (South American rattlesnake).  
OS Eukaryota; Metazoa; Chordata; Cranialata; Euteleostomi;  
OC Lepidodermata; Squamata; Scleroglossa; Serpentes; Colubroidae;  
OC Viperidae; Crocaltinae; Crocatus.  
OX NCBI\_TaxID=8732;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=VENOM GLAND;  
RA Hayashi M.A.F., Radis-Baptista G., Barbosa S.R., Yamane T., Camargo A.C.M.;  
RT "Crocatus durissus terrificus bradykinin-potentiating peptide and C-type natriuretic peptide precursor isoform2.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF308594; AAL09427.1; -;  
DR InterPro: IPR000663; Natr\_peptide.  
DR Pfam: PF00212; ANP; 1.  
DR PROSITE: PS00263; NATRIURETIC PEPTIDE; UNKNOWN 1.  
SO SEQUENCE 181 AA; 18507 MW; 982E95D38A5FF27 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 8; DB 13; Length 181;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLTALL 1140  
DB 10 GLLTALL 17

## RESULT 33

Q97LC8 PRELIMINARY; PRT; 185 AA.  
AC Q97LC8;  
DT 01-OCT-2001 (TRENBLREL. 18, Created)  
DT 01-OCT-2001 (TRENBLREL. 18, Last sequence update)  
DT 01-OCT-2001 (TRENBLREL. 18, Last annotation update)  
DE Predicted membrane protein.  
GN CAC0634.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VGM B-1787;

```

RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007579; AAK78611.1; -.
KM Complete proteome.
SQ SEQUENCE 185 AA; 20003 MW; DEE2B40DD56612CB CRC64;

Query Match 0.7%; Score 8; DB 16; Length 185;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 636 AALLSSR 643
DB 54 AALLSSR 61

RESULT 34
O9YBV7 PRELIMINARY; PRT; 240 AA.
AC O9YBV7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein APL1493.
GN APL1493.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jinno K., Takahashi M., Sekine M., Baba S.-I., Ankaei A., Koeugli H.,
RA Hoshizawa A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kudota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000061; BAA80491.1; -.
DR InterPro; IPR003754; HEM4.
DR Pfam; PF02602; HEM4; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 240 AA; 25442 MW; 14EB211AD0646A7A CRC64;

Query Match 0.7%; Score 8; DB 17; Length 240;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 GYSVSSML 439
DB 107 GYSVSSML 114

RESULT 35
O9RK95 PRELIMINARY; PRT; 258 AA.
AC O9RK95;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative hydrolase.
GN SC00267 OR SCF1.09.
OS Streptomyces coelicolor.
RT Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

```

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)";
RL Nature 417:141-147(2002).
DR EMBL; AL117322; CAB55529.1; -.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR003089; AB hydrolase.
DR Pfam; PF00561; abhydrolase; 1.
DR PRINTS; PR00111; ABHYDROLASE.
KM Hydrolase.
SQ SEQUENCE 258 AA; 27454 MW; 4E9B7CFECA5802C5 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 258;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 ELPAALKA 291
DB 156 ELPAALKA 163

RESULT 36
O9PW56 PRELIMINARY; PRT; 265 AA.
AC O9PW56;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Bradykinin-potentiating/C-type natriuretic peptide.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylia; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Hayaishi M.A.F., Murbach A.F., Camargo A.C.M.;
RT "The precursor of C-type natriuretic peptide of snake brain contains
RT angiotensin converting enzyme inhibitors, specific for the C-catalytic

```

RT site."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF171670; AAD51326.2; -  
 DR InterPro: IPR000663; Nact\_peptide.  
 DR Pfam: PF00212; ANP; 2  
 DR PRINTS: PRO0710; NATPEPTIDES.  
 DR SMART: SM00183; NAT\_PEP; 2.  
 DR PROSITE: PS00263; NATRIURETIC PEPTIDE; 2.  
 SQ SEQUENCE 265 AA; 2763 MW; 8E99AEC976CCD439 CRC64;

Query Match 0.7%; Score 8; DB 13; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLLLALL 1140  
 |||||  
 DB 10 GLLLLALL 17

## RESULT 37

08CG91 PRELIMINARY; PRT; 265 AA.  
 AC 08CG91;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Bradykinin-potentiating/c-type natriuretic protein.  
 OS Bothrops insularis (Island jararaca) (Queimada jararaca).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidae;  
 OC Vipéridae; Crotalinae; Bothrops.  
 OX NCBI\_TaxID=8723;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=VENOM GLAND;  
 RA Unpublished de-Azevedo I.L.M., Ho P.L.;  
 RT "A survey of gene expression and diversity in the venom glands of the  
 RT pitviper Bothrops insularis through the generation of Expressed  
 RT Sequence Tags (ESTs).";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF490531; AAM09690.1; -  
 SQ SEQUENCE 265 AA; 2763 MW; 0EAE1408B42358BE CRC64;

Query Match 0.7%; Score 8; DB 13; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLLLALL 1140  
 |||||  
 DB 10 GLLLLALL 17

## RESULT 38

09U168 PRELIMINARY; PRT; 269 AA.  
 AC 09U168;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE Hypothetical 28.8 kDa protein.  
 GN L377.08.  
 OS Leishmania major.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIDLIN;  
 RA Masny D., Furnelle B., Goffeau A., Ivens A.C., Murphy L., Quail M.,  
 RA Rajandream M.A., Barrell B.G.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=FRIDLIN;

RX MEDLINE=98146435; PubMed=9477341;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 RA Smith D.F.;  
 RT "A physical map of the Leishmania major Friedlin genome.";  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL: AL136326; CAB65934.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 269 AA; 28805 MW; F35F62D65882141A CRC64;

Query Match 0.7%; Score 8; DB 5; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 600 AAASMPHA 607  
 |||||  
 DB 75 AAASMPHA 82

## RESULT 39

053905 PRELIMINARY; PRT; 281 AA.  
 AC 053905;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ACTVA 3 protein.  
 GN ACTVA 3.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=92114870; PubMed=1766437;  
 RA Caballero J.L., Martinez E., Malpartida F., Hopwood D.A.;  
 RT "Organization and functions of the actva region of the actinorhodin  
 RT biosynthetic gene cluster of Streptomyces coelicolor.";  
 RL Mol. Gen. Genet. 230:401-412(1991).  
 DR EMBL: X58833; CAA41639.1; -  
 SQ SEQUENCE 281 AA; 29952 MW; 55EC061CE3C2EFB3 CRC64;

Query Match 0.7%; Score 8; DB 2; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 708 TAGARAAF 715  
 |||||  
 DB 89 TAGARAAF 96

## RESULT 40

0931Z2 PRELIMINARY; PRT; 281 AA.  
 AC 0931Z2;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein actva3.  
 GN ACTVA3 OR SCOS078 OR SCBAC28G1.04.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser H., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; A593842; CAC44191.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 281 AA; 30028 MW; DFC4A496C340E1B3 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 8; DB:16; Length 281;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 708 TAGARAA 715  
 |||||  
 Db 89 TAGARAA 96

## RESULT 41

0930X2 PRELIMINARY; PRT; 287 AA.  
 ID 0930X2;  
 AC 0930X2;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Lycopen elongase.  
 GN CRTB.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
 OC Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD23;  
 RA Krubasik P., Kobayashi M., Sandmann G.;  
 RT "Expression and functional analysis of a gene cluster involved in the  
 RT synthesis of decaprenoxanthin reveals the mechanisms for C50  
 RT carotenoid formation.";  
 RL Eur J Biochem. 286:3703-3708(2001).  
 DR EMBL; AF159510; AKK64302.1; -;  
 DR InterPro; IPR000537; UblA.  
 DR Pfam; PF01040; UblA; 1.  
 SQ SEQUENCE 287 AA; 31674 MW; 3FD29E1C6DDASFOA CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 8; DB:2; Length 287;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 ILLSRPI 645  
 |||||  
 Db 8 ILLSRPI 15

## RESULT 42

09XCM9 PRELIMINARY; PRT; 294 AA.  
 ID 09XCM9;  
 AC 09XCM9;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE NoSF.  
 GN NOSF.  
 OS Rhodobacter sphaeroides f. sp. denitrificans.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Rhodobacter.  
 OX NCBI\_TaxID=39723;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL106;  
 RX MEDLINE=99429850; PubMed=10498715;  
 RA Sabaty M., Schwintner C., Cahors S., Richard P., Vermeglio A.;

RT "Nitrite and nitrous oxide reductase regulation by nitrogen oxides in  
 RT Rhodobacter sphaeroides f. sp. denitrificans Il106.";  
 RL J. Bacteriol. 181:6028-6032(1999).  
 DR EMBL; AF15260; AAD3475.1; -;  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transport.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transport; 1.  
 DR SMART; SMO0382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 KW ATP-binding.  
 SQ SEQUENCE 294 AA; 31061 MW; C76B100453535E5 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 8; DB:2; Length 294;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 GAAILLS 642  
 |||||  
 Db 176 GAAILLS 183

## RESULT 43

084105 PRELIMINARY; PRT; 298 AA.  
 ID 084105;  
 AC 084105;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE HAD superfamily hydrolase/phosphatase.  
 GN CT103.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=D/OW-3/CX;  
 RX MEDLINE=9900809; PubMed=9784136;  
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis.";  
 RL Science 282:754-759(1998).  
 DR EMBL; AE001284; AAC67694.1; -;  
 KW Hydrolase; Complete proteome.  
 SQ SEQUENCE 298 AA; 34254 MW; 9F2ADCCFDF67A74D CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 8; DB:16; Length 298;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 PEARLIV 272  
 |||||  
 Db 141 PEARLIV 148

## RESULT 44

09R9K7 PRELIMINARY; PRT; 299 AA.  
 ID 09R9K7;  
 AC 09R9K7;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE NoSF protein.  
 GN NOSF.  
 OS Paracoccus denitrificans.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Paracoccus.  
 OX NCBI\_TaxID=266;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PD1222;

RX MEDLINE=20416234; PubMed=10960107;  
 RA Saudere N.F.W., Hrnberg J.J., Reijnders W.N.M., Westerhoff H.V.,  
 RA de Vries S., van Spanning R.J.M.;  
 RT "The NosX and Nix proteins of *Paracoccus denitrificans* are functional  
 RT homologues: Their role in maturation of nitrous oxide reductase";  
 RL J. Bacteriol. 182:5211-5217(2000).  
 DR EMBL; AF010260; CAB53353.1; -  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR Pfam; PF000005; ABC\_tran.1.  
 DR ProDom; PD000006; ABC\_transportr; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 KW ATP-binding.  
 SQ SEQUENCE 299 AA; 31687 MW; 69849D3F85456214 CRC64;

Query Match 0.7%; Score 8; DB 2; Length 299;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 GAALLLS 642  
 DB 179 GAALLLS 186

## RESULT 45

094K03 PRELIMINARY; PRT; 299 AA.

AC 094K03;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Unknown protein.  
 GN A\_IIG002N01.8  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinzaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK370497; AAK43874.1; -  
 SQ SEQUENCE 299 AA; 34028 MW; BB2AB03C4A184D3E CRC64;

Query Match 0.7%; Score 8; DB 10; Length 299;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141  
 DB 44 LLLALLV 51

## RESULT 46

005656 PRELIMINARY; PRT; 309 AA.

AC 005656;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Putative malate oxidoreductase (Fragment).  
 GN MAOX  
 OS *Bacillus cereus*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1396;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 10987;  
 RA Koltoe A.B.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 10987;  
 RX MEDLINE=99231848; PubMed=10217496;  
 RA Oksend O.A., Hegna I., Lindback T., Rishovd A.L., Koltoe A.B.;  
 RT "Genome organisation is not conserved between *Bacillus cereus* and  
 RT *Bacillus subtilis*."  
 RL Microbiology 145:621-631(1999).  
 DR EMBL; Y09212; CAA70412.1; -  
 DR InterPro; IPR001891; Malic\_oxred.  
 DR Pfam; PF00390; malic; 1.  
 DR PRINTS; PR00072; MALOXRDASE.  
 DR PROSITE; PS00331; MALIC\_ENZYMES; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 309 309  
 SQ SEQUENCE 309 AA; 34451 MW; 5FF0C7A9BF0470EF CRC64;

Query Match 0.7%; Score 8; DB 2; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 LVVVTDGE 278  
 DB 170 LVVVTDGE 177

## RESULT 47

08TE55 PRELIMINARY; PRT; 348 AA.

AC 08TE55;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE FLJ00114 protein (Fragment).  
 GN FLJ00114.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPLEEN;  
 RA Chata O., Nagase T., Kikuno R., Okumura K.;  
 RT "The nucleotide sequence of a long CDNA clone isolated from human  
 RT spleen."  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK074047; BAB84873.1; -  
 FT NON\_TER 1 1  
 SQ SEQUENCE 348 AA; 38941 MW; 5A698D66A1C9992 CRC64;

Query Match 0.7%; Score 8; DB 4; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
 DB 300 GGLLLAL 307

## RESULT 48

09M146 PRELIMINARY; PRT; 360 AA.

AC 09M146;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE Hypothetical 41.1 kDa protein.  
 GN ATAG01220.



OC Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL161491; CAB80931.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 360 AA; 41120 MW; A7E4CBAB9369D931 CRC64;  
 Query Match 0.7%; Score 8; DB 10; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1134 LLLALLV 1141  
 Db 44 LLLALLV 51  
 RESULT 49  
 016339  
 ID 016339 PRELIMINARY; PRT; 383 AA.  
 AC 016339;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE F59P6.2 protein.  
 GN F59P6.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=BRISTOL N2;  
 RC MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,  
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=BRISTOL N2;  
 RC Bradshaw H., Graves T.;  
 RL "The sequence of C. elegans cosmid F59D6.";  
 RT Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=BRISTOL N2;  
 RC Waterston R.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF014335; AAB65877.1; -  
 DR HSSP; P00794; 4CWS.  
 DR InterPro; IPR001461; AspproteaseA1.  
 DR Pfam; PF00026; asp; 1.  
 SQ SEQUENCE 383 AA; 42004 MW; D9DE356D3AB7401E CRC64;

Query Match 0.7%; Score 8; DB 5; Length 383;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1136 LLLALLVFC 1143  
 Db 8 LLLALLVFC 15  
 RESULT 50  
 024416  
 ID 024416 PRELIMINARY; PRT; 385 AA.  
 AC 024416;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Possible acyltransferase, copy 1.  
 GN L8032.06.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OC NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,  
 RA Rajandream M.A., Barrell B.G.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA MEDLINE=96146435; PubMed=9477241;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 RA Smith D.F.;  
 RT "A physical map of the Leishmania major Friedlin genome.";  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL; AL139794; CAC22648.1; -  
 DR InterPro; IPR002123; Acyltransferase.  
 DR Pfam; PF01553; Acyltransferase; 1.  
 KW Acyltransferase; Transferase.  
 SQ SEQUENCE 385 AA; 43595 MW; D4BBD5E7E8B9791 CRC64;  
 Query Match 0.7%; Score 8; DB 5; Length 385;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 716 DSGGRLS 723  
 Db 162 DSGGRLS 169  
 RESULT 51  
 024416  
 ID 024416 PRELIMINARY; PRT; 405 AA.  
 AC 024416;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Pectate lyase.  
 GN PL.  
 OS Fragaria ananassa (Strawberry).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustids I; Rosales; Rosaceae; Rosoideae; Fragaria.  
 OC NCBI\_TaxID=3747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=CHANDLER;  
 RC MEDLINE=97435972; PubMed=9290639;  
 RA Medina-Escobar N., Cardenas J., Moyano E., Caballero J.L.,  
 RA Munoz-Bianco J.;  
 RT "Cloning, molecular characterization and expression pattern of a  
 RT strawberry ripening-specific cDNA with sequence homology to pectate

RT lyase from higher plants.";  
 RL Plant Mol. Biol. 34:867-877(1997).  
 DR EMBL: U63550; AAB71208.1; -;  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam: PF00544; pec\_lyase; 1.  
 KW Lyase.  
 SQ SEQUENCE 405 AA; 45744 MW; 563BC01BCA640599 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 405;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141  
 DB 18 LLLALLLV 25

RESULT 52  
 Q94FT6 PRELIMINARY; PRT; 405 AA.  
 AC Q94FT6;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Pectate lyase B (Fragment).  
 OS Fragaria ananassa (Strawberry).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroids I; Rosales; Rosaceae; Rosoideae; Fragaria.  
 OC NCBI\_TaxID=3747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. CHANDLER;  
 RA Munoz-Bianco J., Caballero J.L., Benitez-Burraco A.;  
 RT Ripening-related strawberry pectate-lyase genes.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF33024; AAK66160.1; -;  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR Pfam: PF00544; pec\_lyase; 1.  
 KW Lyase.  
 FT NON\_TER 405 405.  
 SQ SEQUENCE 405 AA; 45654 MW; DB169E2E59BF8300 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 405;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141  
 DB 18 LLLALLLV 25

RESULT 53  
 Q915T3 PRELIMINARY; PRT; 415 AA.  
 ID Q915T3;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE Probable permease of ABC transporter.  
 GN PA0605.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RA MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004497; AAG03994.1; -;  
 DR InterPro: IPR000515; BPD\_transp.  
 DR Pfam: PF00528; BPD\_transp; 1.  
 DR PROSITE: PS00402; BPD\_TRANSP\_INN\_MEMBER; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 415 AA; 45925 MW; A0F3425BA8E60468 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLLA 1138  
 DB 388 LGGLLLLA 395

RESULT 54  
 Q8YGR5 PRELIMINARY; PRT; 433 AA.  
 ID Q8YGR5;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein A1373754.  
 GN ALR3754.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OC NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL: AP003594; BAB75453.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 433 AA; 48889 MW; 5991F41A9C80C0B4 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 433;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 982 VSGLIISA 989  
 DB 247 VSGLIISA 254

RESULT 55  
 O04611 PRELIMINARY; PRT; 444 AA.  
 ID O04611;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE A\_IG002N01.8 protein.  
 GN A\_IG002N01.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;  
 RA Schect P., Maggi L.;  
 RT "The sequence of A. thaliana IG002N01."  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Wash-U;  
 RT "The A. thaliana Genome Sequencing Project."  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Waterston R.;  
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF007269; AAB61018.1;  
 SO SEQUENCE 444 AA; 50340 MW; A670696CA0052DF4 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 444;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141  
 Db 115 LLLALLLV 122

RESULT 56  
 ID 08XZT4 PRELIMINARY; PRT; 473 AA.  
 AC 08XZT4;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Probable replicative DNA helicase protein (EC 3.6.1.-).  
 GN DNAB OR RSC1311 OR RS02833.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMT1000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brotier P., Camus J.C., Cattoilco L.,  
 RA Chandler M., Choisme N., Claudel-Renaud C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie W., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Sigler P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum."  
 RL Nature 415:497-502 (2002).  
 DR EMBL; AL646063; CAD15013.1;  
 DR InterPro; IPR003593; AAA\_Artpase.  
 DR InterPro; IPR001198; DnaB\_helicase.  
 DR Pfam; PF00772; DnaB; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00665; DnaB; 1.  
 DR Hydrolase; Complete proteome.  
 KM SEQUENCE 473 AA; 52293 MW; 155F5FC98B5434E6 CRC64;  
 SO SEQUENCE 473 AA; 52293 MW; 155F5FC98B5434E6 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 473;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1129 SVLGILL 1136  
 Db 34 SVLGILL 41

RESULT 57  
 Q9REL8 PRELIMINARY; PRT; 491 AA.

AC Q9REL8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Response regulator HupR.  
 GN HUPR.  
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Rhodobacter.  
 OX NCBI\_TaxID=1063;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=601;  
 RA Colbeau A., Chabert J., Wu J.-Q., Vignais P.M.;  
 RT "Hup gene cluster from Rhodobacter sphaeroides."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTRAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING DOMAIN.  
 CC EMBL; AF214145; AAF19999.2;  
 DR HSP; P41789; INTR.  
 DR InterPro; IPR002197; HTH\_Fis.  
 DR InterPro; IPR001789; Response\_reg.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR Pfam; PF02954; HTH\_8; 1.  
 DR Pfam; PF00072; response\_reg; 1.  
 DR Pfam; PF00158; Sigma54\_activat; 1.  
 DR PRINTS; PR01590; HTHPRIS.  
 DR ProDom; PD000039; Response\_reg; 1.  
 DR SMART; SM00448; REC; 1.  
 DR TIGRFAMs; TIGR01199; HTH\_fis; 1.  
 DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; UNKNOWN\_1.  
 DR PROSITE; PS00445; SIGMA54\_INTERACT\_4; 1.  
 KW ATP-binding; DNA-binding; Phosphorylation; Sensory transduction;  
 KW Transcription regulation.  
 SO SEQUENCE 491 AA; 53928 MW; 09D348AE76A328 CRC64;

Query Match 0.7%; Score 8; DB 2; Length 491;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 RMALEDEF 420  
 Db 20 RMALEDEF 27

RESULT 58  
 ID Q9K559 PRELIMINARY; PRT; 491 AA.  
 AC Q9K559;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative hydrolase transcriptional regulator (HupR protein).  
 GN HUPR.  
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Rhodobacter.  
 OX NCBI\_TaxID=1063;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=601;  
 RA Xu D., Liang J., Wu Y.;  
 RT "The study of the hupR gene in Rhodobacter sphaeroides 601."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=601;  
 RA Xu D.O., Wu Y.O.;  
 RT "Cloning of the hup gene cluster from Rhodobacter sphaeroides 601."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=601;

RA Xu D.Q., Wu Y.Q.;  
 RT "Cloning, sequencing and function of the huPr gene from Rhodobacter  
 RL sphaeroides 601.";  
 CC Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING  
 DOMAIN.

DR EMBL; AJ243734; CAC01253.1; -;  
 DR EMBL; AJ315650; CAC42238.1; -;  
 DR HSSP; P41789; INTR.

DR InterPro; IPR002197; HTH\_Fis.  
 DR InterPro; IPR001789; Response\_reg.

DR InterPro; IPR002078; Sig54\_interact.

DR Pfam; PF02954; HTH\_8; 1.  
 DR Pfam; PF00072; response\_reg; 1.  
 DR Pfam; PF00158; Sigma54\_activat; 1.

DR PRINTS; PR01590; HTHFIS.  
 DR ProDom; PD000039; Response\_reg; 1.

DR SMART; SM00448; REC; 1.  
 DR TIGRFAMs; TIGR01199; HTH\_Fis; 1.

DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; UNKNOWN\_1.  
 DR PROSITE; PS00645; SIGMA54\_INTERACT\_4; 1.

KW ATP-binding; DNA-binding; Phosphorylation; Sensory transduction;  
 KW Transcription regulation.

SEQUENCE 491 AA; 54031 MW; 348616E053BBD5AE CRC64;

Query Match 0.7%; Score 8; DB 2; Length 491;  
 Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 LALLALLLV 420  
 DB 20 LALLLEDEF 27

RESULT 59  
 ID Q9WUD0 PRELIMINARY; PRT; 491 AA.

AC Q9WUD0;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Cytochrome P450 2B10 related protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C; TISSUE=LIVER;  
 RA MEDLINE=99240492; PubMed=10222226;

RT "Isolation of a cyp2b10-like cDNA and of a clone derived from a  
 RT cyp2b10-like pseudogene.";

RL Biochem. Biophys. Res. Commun. 258:11-16(1999).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; AF128849; AAD28466.1; -;  
 DR HSSP; P00179; IDT6.

DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.

DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.

KW Heme; Monooxygenase; Oxidoreductase.

SEQUENCE 491 AA; 55870 MW; A1338E108D77025E CRC64;

Query Match 0.7%; Score 8; DB 11; Length 491;  
 Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LALLALLLV 1141  
 DB 6 LALLALLLV 13

RESULT 60

ID Q64584 PRELIMINARY; PRT; 491 AA.

AC Q64584;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Rat cytochrome P-450b (phenobarbital-inducible).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]  
 RP SEQUENCE FROM N.A.

RC MEDLINE=85234490; PubMed=2989270;  
 RA Suwa Y., Mizukami Y., Sogawa K., Fujii-Kuriyama Y.;

RT "Gene structure of a major form of phenobarbital-inducible cytochrome  
 RT P-450 in rat liver.";

RL J. Biol. Chem. 260:7980-7984(1985).

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; L00320; AAA41046.1; JOINED.

DR EMBL; M1251; AAA41046.1; JOINED.

DR EMBL; L00313; AAA41046.1; JOINED.

DR EMBL; L00314; AAA41046.1; JOINED.

DR EMBL; L00315; AAA41046.1; JOINED.

DR EMBL; L00316; AAA41046.1; JOINED.

DR EMBL; L00317; AAA41046.1; JOINED.

DR EMBL; L00318; AAA41046.1; JOINED.

DR EMBL; L00319; AAA41046.1; JOINED.

DR HSSP; P00179; IDT6.

DR InterPro; IPR001128; Cytochrome\_P450.

DR Pfam; PF00067; P450; 1.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.

KW Heme; Monooxygenase; Oxidoreductase.

SEQUENCE 491 AA; 56047 MW; E9144D72E522F0C CRC64;

Query Match 0.7%; Score 8; DB 11; Length 491;  
 Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LALLALLLV 1141  
 DB 6 LALLALLLV 13

RESULT 61  
 ID Q73804 PRELIMINARY; PRT; 497 AA.

AC Q73804;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Hypothetical 56.4 kDa protein.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Takifugu.

OX NCBI\_TaxID=31033;

RN [1]  
 RP SEQUENCE FROM N.A.

RC MEDLINE=99177347; PubMed=10077531;  
 RA Gellner K., Brenner S.;

RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu  
 RT rubripes.";

RL Genome Res. 9:251-258(1999).  
 DR EMBL; AF056116; AAC34385.1; -;  
 KW Hypothetical protein.

SEQUENCE 497 AA; 56376 MW; 802743E93EB93E64 CRC64;

Query Match 0.7%; Score 8; DB 13; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLLLLV 1141  
 |||||  
 169 LLLLLLV 176

## RESULT 62

O9AS41

ID O9AS41 PRELIMINARY; PRT; 518 AA.

AC O9AS41:  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Putative cytochrome P450.  
 GN P0416G1.1 OR B111C09.13.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 ON NCBI\_TaxID=4530;  
 RX [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone: P0416G1.1";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC  
 clone: B111C09.1";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; AP002968; BAB39252.1; -;  
 DR EMBL; AP003204; BAB61166.1; -;  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.

DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 DR Heme; Monooxygenase; Oxidoreductase.  
 KW SEQUENCE 518 AA; 57504 MW; 5A58E8246CC7ABC9 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 518;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLLLLV 1141  
 |||||  
 8 LLLLLLV 15

## RESULT 63

O8WVY3

ID O8WVY3 PRELIMINARY; PRT; 573 AA.

AC O8WVY3:  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 64.3 kDa protein.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Primates; Catarrhini; Homiinae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=CERVIX;  
 RA Strausberg R.;

RU Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC013741; AAH13741.1; -;  
 DR InterPro; IPR001909; KRAB.

DR InterPro; IPR000822; Znf\_C2H2.  
 DR Pfam; PF00096; ZF-C2H2; 12.  
 DR PRINTS; PR00048; ZINC\_FINGER.  
 DR ProDom; PD000003; Znf\_C2H2; 10.  
 DR SMART; SM00355; Znf\_C2H2; 12.  
 DR PROSITE; PS50805; KRAB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_12.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 12.  
 KW Hypothetical protein.  
 SQ SEQUENCE 573 AA; 64291 MW; 446498FAA8BE73E7 CRC64;

Query Match 0.7%; Score 8; DB 4; Length 573;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 898 FSCSSLLS 905  
 |||||  
 DB 270 FSCSSLLS 277

## RESULT 64

O8R6N2

ID O8R6N2 PRELIMINARY; PRT; 577 AA.

AC O8R6N2:  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE ABC-type multidrug/protein/lipid transport system, ATPase  
 component.  
 GN MDL812 OR TTE2769.

OS Thermomicrobacter tengcongensis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Thermomicrobacteriales; Thermomicrobacteriaceae; Thermomicrobacter.

ON NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB4T / JCM11007;  
 RX MEDLINE=21992816; PubMed=11997336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome";  
 RL Genome Res. 12:689-700(2002).  
 DR EMBL; AE013214; AAM25873.1; -;  
 KW Complete proteome.

SQ SEQUENCE 577 AA; 66097 MW; BA6642DF1FBE7545 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 577;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 226 EEVYRAAK 233  
 |||||  
 DB 429 EEVYRAAK 436

## RESULT 65

O8TRM8

ID O8TRM8 PRELIMINARY; PRT; 596 AA.

AC O8TRM8:  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE F420-nonreducing hydrogenase.

GN VHTA OR MA1147.  
 OS Methanosaarcina acetivorans.  
 OC Archaea; Euryarchaeota; Methanococci; Methanosaarcinales;  
 OC Methanosaarcinaceae; Methanosaarcina.  
 ON NCBI\_TaxID=2214;  
 RX [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=1193238;

```

RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeRellano K., Johnson R.,
RA Linton L., McKernan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umagay L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.,
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010781; AAM04568.1; -.
KW Complete proteome.
SQ SEQUENCE 596 AA; 65478 MW; 3780AFC2448F9D5B CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 85;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 44 SVLQHWGG 51
Db 187 SVLQHWGG 194

RESULT 66
ID Q39775 PRELIMINARY; PRT; 607 AA.
AC Q39775;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Legumin, 11S globulin.
OS Gnetum gnetum.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Gnepotphyta; Gnecopsida; Gnecatales; Gnecaceae; Gnetum.
OX NCBI_Taxid=3382;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RX MEDLINE=98440765; PubMed=9767693;
RA Shucov A.D., Braun H., Chesnokov Y.V., Horstmann C., Kakhovskaya I.A.,
RA Baumelein H.;
RT "Sequence peculiarity of Gnetalean legumin-like seed storage
RT proteins."
RL J. Mol. Evol. 47:486-492(1998).
DR EMBL: Z50779; CAA90642.1; -.
DR InterPro: IPR000459; Seedstore 11S.
DR Pfam: PF00190; Seedstore 11S; I.
DR PRINTS: PR00439; 11SGLOBULIN.
SQ SEQUENCE 607 AA; 68910 MW; 7AD6DDC604FB3A60 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 86;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1134 LLLALLLV 1141
Db 10 LLLALLLV 17

RESULT 67
ID Q31615 PRELIMINARY; PRT; 614 AA.
AC Q31615;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE YUBQ protein.
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OS Gnathostomata; Actinopterygii; Cyprinodontiformes;
OS Bacillus subtilis.

```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunat F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azeredo V., Bortner M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bortner L., Brans A., Braun M., Bignelli S.C., Bron S.,
RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Comercon I.F., Cummings N.J., Daniel R.A.,
RA Deutz F., Devine K.M., Distenhof A., Ehrlich S.D., Emerson P.T.,
RA Eutenei K.D., Ewington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaetr-Blanchard N., Klein C.,
RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,
RA Prescan E., Puig P., Purnelle D., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowaka A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Mamut R., Medler E., Medler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunat F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z99110; CAB13021.1; -.
DR InterPro: IPR000676; Nah_Eckng.
DR InterPro: IPR000309; TrkA_Kupfke.
DR InterPro: IPR003148; TrkA_N.
DR Pfam: PF00939; Na_H_Exchange; 1.
DR Pfam: PF02080; TrkA-C; 1.
DR Pfam: PF02254; TrkA-N; 1.
KW Complete proteome.
SQ SEQUENCE 614 AA; 67467 MW; EAAAE1850799163B CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 87;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1134 LLLALLLV 1141
Db 313 LLLALLLV 320

RESULT 68
ID Q91570 PRELIMINARY; PRT; 622 AA.
AC Q91570;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Integrin alpha5tr subunit (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae.

```

```

OC Xenopodinae; Xenopus.
OK NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=95344594; PubMed=7619730;
RA Joos T.O., Whitlaker C.A., Meng F., Desimone D.W., Gnan V., Hansen P.;
RT "Integrin alpha 5 during early development of Xenopus laevis.";
RL Mech. Dev. 50:187-199(1995).
DR EMBL; U12646; AAA96667.1; -.
DR HSSP; P06756; 1JY2.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 1.
FT NON TER
SQ SEQUENCE 622 AA; 69003 MW; 78090D59A6B3660C CRC64;

Query Match 0.7%; Score 8; DB 13; Length 622;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLTLALL 1140
Db 604 GLLTLALL 611

RESULT 69
017795 PRELIMINARY; PRT; 634 AA.
AC 017795;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 68.2 kDa protein.
GN C07G1.4
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Hawkins J.;
RT "The sequence of C. elegans cosmid C07G1.";
RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U58751; AAB00657.2; -.
DR InterPro; IPR000095; PARbox/Rhobndng.
DR InterPro; IPR000697; Ranbp1_WASP.
DR InterPro; IPR001960; WH1.
DR InterPro; IPR003124; WH2.
DR Pfam; PF00786; PBD; 2.
DR Pfam; PF00568; WH1; 2.
DR Pfam; PF02205; WH2; 4.
DR PRINTS; PR01582; KV33CHANNEL.
DR SMART; SM00285; PBD; 2.
DR SMART; SM00461; WH1; 2.
DR SMART; SM00246; WH2; 2.
DR PROSITE; PSS0108; GBD; 2.

```

```

KW Hypothetical protein.
SQ SEQUENCE 634 AA; 68201 MW; C9E489254C070DD0 CRC64;

Query Match 0.7%; Score 8; DB 5; Length 634;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 EAVCLTAA 679
Db 48 EAVCLTAA 55

RESULT 70
09CWM6 PRELIMINARY; PRT; 636 AA.
ID 09CWM6
AC 09CWM6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein PM0698.
GN PM0698.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006106; AAK02782.1; -.
DR InterPro; IPR001091; CNA_MetTransf.
DR InterPro; IPR002295; D21N6_mltfrase.
DR InterPro; IPR002941; N6/N4_Mtase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF01555; N6_N4_Mtase; 1.
DR PRINTS; PR00506; D21N6MTFRASE.
DR PRINTS; PR00508; S21N4MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 636 AA; 72597 MW; D3A20817BCE8A6A6A CRC64;

Query Match 0.7%; Score 8; DB 16; Length 636;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 969 KTLRLVQN 976
Db 343 KTLRLVQN 350

RESULT 71
070491 PRELIMINARY; PRT; 670 AA.
ID 070491
AC 070491;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE Retinoic acid-responsive protein.
GN STRA6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95377533; PubMed=7649373;
RA Boulliet P., Oulad-Abdelghani M., Vicaire S., Garnier J.M.,
Schubbaur B., Dolle P., Chambon P.;
RT "Efficient cloning of cDNAs of retinoic acid-responsive genes in P19
embryonal carcinoma cells and characterization of a novel mouse gene,

```

RT Stral (mouse LERK-2/Ed192).";  
 RL Dev. Biol. 170:420-433 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97346723; PubMed=92031140;  
 RA Bouillet P., Sapin V., Chazaud C., Messadegq N., Decimo D., Dolle P.,  
 RA Chambon P.;  
 RA "Developmental expression pattern of Stra6, a retinoic acid-responsive  
 RT gene encoding a new type of membrane protein.";  
 RL Mech. Dev. 63:173-186 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Bouillet P.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF062476; AAC16016.1; -  
 DR MGD; MGI:107742; Stra6  
 SQ SEQUENCE 670 AA; 73774 MW; 6AC669F3EE16A4F7 CRC64;

Query Match 0.7%; Score 8; DB 11; Length 670;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141  
 DB 64 LLLALLV 71

## RESULT 72

ID 082985 PRELIMINARY; PRT; 685 AA.  
 AC 082985;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Ferrichrome transport protein PhnB.  
 GN STY0221.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
 RA Krogsh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852 (2001).  
 DR EMBL; AL627265; CAD01355.1; -  
 DR InterPro; IPR000522; Feccd.  
 DR Pfam; PF01032; Feccd; 2.  
 DR ProDom; PD001557; Feccd; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 685 AA; 73148 MW; F72A8AE03556B4D CRC64;

Query Match 0.7%; Score 8; DB 16; Length 685;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141  
 DB 36 LLLALLV 43

## RESULT 73

Q96FV1

ID Q96FV1 PRELIMINARY; PRT; 704 AA.  
 AC Q96FV1;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE KIAA1937 protein (Fragment).  
 GN KIAA1937.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRN;  
 RX MEDLINE=21456161; PubMed=11572484;  
 RA Nagase T., Kikuno R., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XXI.  
 RT large proteins.";  
 RT DNA Res. 8:179-187 (2001).  
 DR EMBL; AB067524; BAB67830.1; -  
 FT NON\_TER  
 SQ SEQUENCE 704 AA; 82013 MW; AFG8E675A2CC8BEC CRC64;

Query Match 0.7%; Score 8; DB 4; Length 704;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1130 VLGGLLL 1137  
 DB 1 VLGGLLL 8

## RESULT 74

ID Q91N19 PRELIMINARY; PRT; 739 AA.  
 AC Q91N19;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
 DE VP3.  
 OS Kadipiro virus.  
 OC Viruses; dsRNA viruses; Reoviridae; Colitivirus.  
 OX NCBI\_TaxID=104580;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TKT-7075;  
 RX MEDLINE=20273996; PubMed=10811934;  
 RA Attoui H., Billoir F., Biagini P., de Micco P., de Lamballerie X.;  
 RT "Complete sequence determination and genetic analysis of Banna virus  
 RT and Kadipiro virus: proposal for assignment to a new genus  
 RT (Seadornavirus) within the family Reoviridae.";  
 RL J. Gen. Virol. 81:1507-1515 (2000).  
 DR EMBL; AF134510; AAF78851.1; -  
 SQ SEQUENCE 739 AA; 85527 MW; A651D16CC33A5709 CRC64;

Query Match 0.7%; Score 8; DB 12; Length 739;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1125 ILIGSVLG 1132  
 DB 638 ILIGSVLG 645

## RESULT 75

ID Q9CAD3 PRELIMINARY; PRT; 767 AA.  
 AC Q9CAD3;  
 DT 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Putative UDP-N-acetylmutamoylalanyl-D-glutamate--2, 6-diaminopimelate



DE ligase.  
GN F24D7.13.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxId=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=21016719; PubMed=1130712;  
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.U., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luvos J.S., Maiti R., Marziani A., Miltischer J., Miranda M., Nguyen M., Nietman W.C., Osborne B.I., Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D., Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana."  
RT Nature 408:816-820(2000).  
RL Nature 408:816-820(2000).  
DR EMBL; AC011622; MAG52413.1; -.  
DR InterPro; IPR000713; Mur\_ligase.  
DR InterPro; IPR004101; Mur\_ligase\_C.  
DR Pfam; PF01225; Mur\_ligase; 1.  
DR Pfam; PF02875; Mur\_ligase\_C; 1.  
SQ SEQUENCE 767 AA; 85022 MW; 93887DF10F50C748 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 767;  
Best Local Similarity 100.0%; Pred.No.1.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 AVGAPLED 575  
|||||||  
Db 559 AVGAPLED 566

Search completed: July 16, 2003, 08:33:40  
Job time : 58 secs

**THIS PAGE BLANK (USPTO)**





```

Db      161 MSEVOTFLRLVGLFIDPEIQIUGLVQYGSPPVHEWSLGDFTKEEVVRAAKNLSRREG 240
OY      241 RETKTAQAIWVACTEGFSQSHGSRPEARLLVVTVDGSHDEBELPALAKACEAGRVTYR 300
Db      241 RETKTAQAIWVACTEGFSQSHGSRPEARLLVVTVDGSHDEBELPALAKACEAGRVTYR 300
OY      301 GIAVLGHLRRQRDSSFLREIRTIASDPDERFFNNVTDEALTDIIVALGDRIFGLEGS 360
Db      301 GIAVLGHLRRQRDSSFLREIRTIASDPDERFFNNVTDEALTDIIVALGDRIFGLEGS 360
OY      361 HAENSSSFGLEMSQIGFSTHRLKDGILFGMGA YOMGSVLMLEGGHLLPFRMMLDEDF 420
Db      361 HAENSSSFGLEMSQIGFSTHRLKDGILFGMGA YOMGSVLMLEGGHLLPFRMMLDEDF 420
OY      421 PPALONHAAYLGSYSVSSMLRGARLFLSGARFRRHKGVIAPQKXKDAVVAOSLQGE 480
Db      421 PPALONHAAYLGSYSVSSMLRGARLFLSGARFRRHKGVIAPQKXKDAVVAOSLQGE 480
OY      481 QIGSYFGSELCPDIDRDGTTDVLLVAAPMFLGPONKKEGRVYVYLVGQOSILLTQGTIQ 540
Db      481 QIGSYFGSELCPDIDRDGTTDVLLVAAPMFLGPONKKEGRVYVYLVGQOSILLTQGTIQ 540
OY      541 PEPPODAFFGFAMGALPDINODGFADVAAGAPLEBHGQALYLYHGTSQVRRPDAQI 600
Db      541 PEPPODAFFGFAMGALPDINODGFADVAAGAPLEBHGQALYLYHGTSQVRRPDAQI 600
OY      601 AASMHALSYFGSRVDGRDLDDDLVDVAAGAGAILLSRPIVHLTPSLVETPOAIS 660
Db      601 AASMHALSYFGSRVDGRDLDDDLVDVAAGAGAILLSRPIVHLTPSLVETPOAIS 660
OY      661 VVQDRCRRGQAVCLTALCFQVTSRTPGMDHOFYMRFTASLDEMTAGARAAFDGSGQ 720
Db      661 VVQDRCRRGQAVCLTALCFQVTSRTPGMDHOFYMRFTASLDEMTAGARAAFDGSGQ 720
OY      721 RLSPRLRLSVGNVCEQHFHVLDTSYLRVALTVTFALDNTKPGPVLESGPTSIQ 780
Db      721 RLSPRLRLSVGNVCEQHFHVLDTSYLRVALTVTFALDNTKPGPVLESGPTSIQ 780
OY      781 KLVPSSKOCGPNECVTDVLQVNMDIRSKAPFVVGGRKVLVSTLENRKENAYNT 840
Db      781 KLVPSSKOCGPNECVTDVLQVNMDIRSKAPFVVGGRKVLVSTLENRKENAYNT 840
OY      841 SLSTIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFOTGAKVTFLEFEFSC 900
Db      841 SLSTIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFOTGAKVTFLEFEFSC 900
OY      901 SLSLSQVQKGLTASSDSLENRGTLOENTRQTSAYIQYEPHLLFSSSESTLHRYEHPYGT 960
Db      901 SLSLSQVQKGLTASSDSLENRGTLOENTRQTSAYIQYEPHLLFSSSESTLHRYEHPYGT 960
OY      961 PVGPGEFFKTLRVONLGCYVVGSLISALLPAVHAGGNYFLLSOVITNNASCI VONLT 1020
Db      961 PVGPGEFFKTLRVONLGCYVVGSLISALLPAVHAGGNYFLLSOVITNNASCI VONLT 1020
OY      1021 EPPGPVHPEELQHTNRNLNGSNTQCVVRCHLGOLAKGTEVSGLRLVHNEFFRRARFK 1080
Db      1021 EPPGPVHPEELQHTNRNLNGSNTQCVVRCHLGOLAKGTEVSGLRLVHNEFFRRARFK 1080
OY      1081 SLTVVSTELGTREBSVQLTASRWSESLLEVOTRPLISLWLLIGSVLGGILLALL 1140
Db      1081 SLTVVSTELGTREBSVQLTASRWSESLLEVOTRPLISLWLLIGSVLGGILLALL 1140
OY      1141 VFCLWKLGFFAHKKIPEEKREKLEQ 1167
Db      1141 VFCLWKLGFFAHKKIPEEKREKLEQ 1167

```

```

RESULT 2
AAB64657
ID AAB64657 standard, Protein, 1152 AA.
XX
XX AAB64657;
XX

```

```

DT      22-MAR-2001 (first entry)
XX
DE      Human secreted protein BLAST search protein SEQ ID NO: 167.
XX
KW      Cyclostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;
KW      antiallergic; hepatocytic; antidiabetic; antiinflammatory; antitumor;
KW      vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW      cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW      neurological disease; infection; human; secreted protein.
XX
OS      Homo sapiens.
XX
EN      WO200077197-A1.
XX
PD      21-DEC-2000.
XX
PF      01-JUN-2000; 2000WO-US14934.
XX
PR      11-JUN-1999; 99US-0138599.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PP      (ROSE/) ROSEN C A.
XX
PI      Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR      WPI, 2001-032312/04.
XX
PT      Isolated nucleic acid molecule encoding a human secreted protein is
XX      used in preventing, treating or ameliorating a medical condition -
XX
PS      Disclosure; Page 543-546; 558pp; English.
XX
CC      The invention relates to the isolation of genes AAF32757-F32803 encoding
CC      the human secreted proteins AAB64549-B64594. The sequence is a search
CC      result from a BLASTX homology search. The genes and proteins are useful
CC      for preventing, ameliorating or treating medical conditions, e.g. by
CC      protein or gene therapy. The genes are isolated from a range of human
CC      tissues disclosed in the specification. The nucleic acids, proteins,
CC      antibodies and (ant)agonists are useful in the diagnosis, treatment
CC      and prevention of: (a) cancer; e.g. breast and ovarian cancer, and
CC      other cancers of the adrenal gland, bone, bone marrow, breast,
CC      gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC      disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC      anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC      multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC      cardiovascular disorders such as myocardial ischaemias; (d) wound
CC      healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC      and (f) infectious diseases such as viral, bacterial, fungal and
CC      parasitic infections.
XX
SQ      Sequence 1152 AA;
XX
Query Match 98.7%; Score 1152; DB 22; Length 1152;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1 MELPFVTHLFLPLVFLTGLSCPFNLDEHHPPLFPGPPEABRGYVLQHVGGGORMLVGA 60
Db      1 MELPFVTHLFLPLVFLTGLSCPFNLDEHHPPLFPGPPEABRGYVLQHVGGGORMLVGA 60
OY      61 PWDPSGDRRDYVRCFPGAGAHNAPCAKGLHGDYQLQNGSSHPANMHLGMSLLETGDDGG 120
Db      61 PWDPSGDRRDYVRCFPGAGAHNAPCAKGLHGDYQLQNGSSHPANMHLGMSLLETGDDGG 120
OY      121 FMACAPLMSRAQSSVSSSGICARVDASFOPOGSLAPTAQRCPTVMYDVYVILDDGNSIYP 180
Db      121 FMACAPLMSRAQSSVSSSGICARVDASFOPOGSLAPTAQRCPTVMYDVYVILDDGNSIYP 180
OY      181 MSEVOTFLRLVGLFIDPEIQIUGLVQYGSPPVHEWSLGDFTKEEVVRAAKNLSRREG 240
Db      181 MSEVOTFLRLVGLFIDPEIQIUGLVQYGSPPVHEWSLGDFTKEEVVRAAKNLSRREG 240
OY      241 RETKTAQAIWVACTEGFSQSHGSRPEARLLVVTVDGSHDEBELPALAKACEAGRVTYR 300

```

```

DB      241 RETKTQAIIMVACTEBSQSHGCRPEAARLLVVVTGESHDEBELPAALKACAGAVTRY      300
QY      301 GTAVLGHYRRORDPSPFLREITIASDPERFEFNVTEBAALTDIVDLGDRIFGLS      360
DB      301 GTAVLGHYRRORDPSPFLREITIASDPERFEFNVTEBAALTDIVDLGDRIFGLS      360
QY      361 HANESSFGLIEMSQIGFSTHRLKDGILFGVNGAYDMGSGVLMDEGHRLPFPPMALEDEF      420
DB      361 HANESSFGLIEMSQIGFSTHRLKDGILFGVNGAYDMGSGVLMDEGHRLPFPPMALEDEF      420
QY      421 PALQNHAAVLYGYSVSMILRGGRULFLSGAPFRHRKVIYAFOLKKDQAVRVAOSLQGE      480
DB      421 PALQNHAAVLYGYSVSMILRGGRULFLSGAPFRHRKVIYAFOLKKDQAVRVAOSLQGE      480
QY      481 QIGSYGSELCPDPTDRDGTDTDLVLAAMPFLGPONKERTGVVYVVGOSLTLTGCTLO      540
DB      481 QIGSYGSELCPDPTDRDGTDTDLVLAAMPFLGPONKERTGVVYVVGOSLTLTGCTLO      540
QY      541 PEPPDARFGFAMGALPDINODGFADAVAGAPLEDHOGALYLYHGTOSGVPHPAQRIA      600
DB      541 PEPPDARFGFAMGALPDINODGFADAVAGAPLEDHOGALYLYHGTOSGVPHPAQRIA      600
QY      601 AASMPHALSYFGRSVDGRDLDDGDDLVDAVAGOGAAILLSRPVYHLTPSLEVPQAIS      660
DB      601 AASMPHALSYFGRSVDGRDLDDGDDLVDAVAGOGAAILLSRPVYHLTPSLEVPQAIS      660
QY      661 VVORDRRRGOEAVCTALACFOVTSRTGMRDHQFYMFTASLDEMTGARAAPDQSGQ      720
DB      661 VVORDRRRGOEAVCTALACFOVTSRTGMRDHQFYMFTASLDEMTGARAAPDQSGQ      720
QY      721 RLSRRLRLSVGNVCEQLHFVLDTSYLRPVALTPTPALTDTTTPGVLNEGSPTSIQ      780
DB      721 RLSRRLRLSVGNVCEQLHFVLDTSYLRPVALTPTPALTDTTTPGVLNEGSPTSIQ      780
QY      781 KLVPSKDCGDPNECTDLVLQNMDIRSKRAPFVVRGRRVYVSTLERNKENAYNT      840
DB      781 KLVPSKDCGDPNECTDLVLQNMDIRSKRAPFVVRGRRVYVSTLERNKENAYNT      840
QY      841 SLSIISRLHLASLTPORESPIKVECAAPSAHARLCSGHPFOGAKATFLLPEFSC      900
DB      841 SLSIISRLHLASLTPORESPIKVECAAPSAHARLCSGHPFOGAKATFLLPEFSC      900
QY      901 SLSLQGVFCKLTASSDSLEENGTLOENTAOQTSAYIYEPHLLFSSESTLHRYEVPYGTL      960
DB      901 SLSLQGVFCKLTASSDSLEENGTLOENTAOQTSAYIYEPHLLFSSESTLHRYEVPYGTL      960
QY      961 PVGPGEFETTLRVQNLGCVYVSGLLISALLPVAHAGNVFSLSOVITNNASCIVQNT      1020
DB      961 PVGPGEFETTLRVQNLGCVYVSGLLISALLPVAHAGNVFSLSOVITNNASCIVQNT      1020
QY      1021 EPPGPVHPEELQHTRLNGSNTQCCOVACHLQOLAKTEVSGLLRLVHNEFFPRRAXK      1080
DB      1021 EPPGPVHPEELQHTRLNGSNTQCCOVACHLQOLAKTEVSGLLRLVHNEFFPRRAXK      1080
QY      1081 SLTVVSTFELGTEGSGVQLTEASRWSLLEVVQTRPILISLWILIGSVLGLLLALL      1140
DB      1081 SLTVVSTFELGTEGSGVQLTEASRWSLLEVVQTRPILISLWILIGSVLGLLLALL      1140
QY      1141 VFCIMWLGFFAH      1152
DB      1141 VFCIMWLGFFAH      1152

```

## RESULT 3

AAV32243  
ID AAV32243 standard; Protein; 1132 AA.

AC AAV32243;

DT 15-FEB-2000 (first entry)

XX Human integrin subunit alpha-10 splice variant.

```

XX      XX      Integrin alpha-10; 1Sa10; human; trauma; rheumatoid arthritis;
KW      KW      osteoarthritis; osteoarthritis; cancer; atherosclerosis;
KW      KW      inflammation; therapy; cartilage; chondrocyte; osteoblast;
KW      KW      fibroblast; vaccine; marker; splice variant.
OS      OS      Homo sapiens.
FH      FH      Key
FT      FT      Peptide
FT      FT      Protein
FT      FT      W09951639-A1.
PN      PN      14-OCT-1999.
PD      PD      31-MAR-1999; 99WO-SE00544.
PF      PF      02-APR-1998; 98SE-0001164.
PR      PR      28-JAN-1999; 98SE-0000319.
XX      XX      (ACT1-) ACTIVE BIOTECH AB.
PA      PA      Lundgren-Akerlund E;
PI      PI      WPI: 2000-052639/04.
XX      XX      DR      N-PSDB; AA234720.
XX      XX      PT      New isolated integrin subunit alpha-10, used as a marker or target
PT      PT      molecule for cells during development, regeneration and pathological
PT      PT      conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
PT      PT      inflammation -
XX      XX      Claim 1; Page 43-48; 90pp; English.
PS      PS      This sequence represents a splice variant of novel human
CC      CC      chondrocyte integrin subunit alpha-10 (1Sa10). It is identical to
CC      CC      1Sa10 (see AAV32242) except for deletion of amino acids 975-986. The
CC      CC      invention relates to a recombinant or isolated integrin heterodimer
CC      CC      comprising the alpha10 subunit in association with subunit beta
CC      CC      (especially beta-1). The heterodimer, subunit alpha-10 or splice
CC      CC      variant can be used as a marker or target of all types of cells, e.g.
CC      CC      of chondrocytes, osteoblasts and fibroblasts. They can also be used:
CC      CC      for treating pathological conditions involving 1Sa10, such as
CC      CC      damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
CC      CC      for detecting the formation of cartilage during embryonal
CC      CC      development, physiological or therapeutic repairation of cartilage,
CC      CC      or detecting regeneration of cartilage or chondrocytes during
CC      CC      transplantation of cartilage or chondrocytes; for selection and
CC      CC      analysis or for sorting, isolating or purification of chondrocytes;
CC      CC      and for in vitro studies of differentiation of chondrocytes; and as
CC      CC      a target for anti-adhesive drugs or molecules in tendon, ligament,
CC      CC      skeletal muscle or other tissues where adhesion impairs the function
CC      CC      of the tissue (all claimed). 1Sa10 binding entities can be used to
CC      CC      determine the differentiation-state of cells during embryonic
CC      CC      development, angiogenesis or development of cancer, in pathological
CC      CC      conditions such as rheumatoid arthritis, osteoarthritis or cancer,
CC      CC      in tissue regeneration or in therapeutic and physiological repairation
CC      CC      of cartilage (claimed). A vaccine comprising the integrin
CC      CC      heterodimer or subunit alpha-10 is also claimed. 1Sa10
CC      CC      polynucleotides, vectors, host cells and methods of producing
CC      CC      recombinant 1Sa10 are also claimed.
XX      XX      Sequence      1132 AA;
SQ      SQ      Query Match      83.4%; Score 973; DB 21; Length 1132;
XX      XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      XX      Matches 973; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      QY      1 MELPVTHLFLPLVFLNGLCPFLNLDENHRLFGPPRAEFGYVLQHVGGGQRMVLVGA 60

```

```

Db      1 MELPVTHLFLPLVLTGLGCSFNLDEHHRLFPGPPEAREGYSLVQHVGGQRMMLVGA 60
Qy      61 FWDGSGDRGDVYRCVPGAHNAPCAKGLDYOJGNSSHPAVMHLMGSLLETDDGCG 120
Db      61 FWDGSGDRGDVYRCVPGAHNAPCAKGLDYOJGNSSHPAVMHLMGSLLETDDGCG 120
Qy      121 FMACPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVYIVLDGNSIYP 180
Db      121 FMACPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVYIVLDGNSIYP 180
Qy      181 MSEVOTPLRLVGLKFLIDPEQIOVGLVOYGSPVHEWSLGDPRTEEYVRAAKNLSRREG 240
Db      181 MSEVOTPLRLVGLKFLIDPEQIOVGLVOYGSPVHEWSLGDPRTEEYVRAAKNLSRREG 240
Qy      241 RETKTAQAIMVACTGFSQSHGRPEARALVVVTDGSHDQELPALKACEARVTRY 300
Db      241 RETKTAQAIMVACTGFSQSHGRPEARALVVVTDGSHDQELPALKACEARVTRY 300
Qy      301 GIAVIGHYLRORDSSFLREIRITIASDPDERFFPNVTDEAALTDIVDALGDRIFGLSGS 360
Db      301 GIAVIGHYLRORDSSFLREIRITIASDPDERFFPNVTDEAALTDIVDALGDRIFGLSGS 360
Qy      361 HAENSSPGLSESOIGFSTHRLKDGILFGMVGAYDMGSLVLMLEGHRLPPRMALDEBF 420
Db      361 HAENSSPGLSESOIGFSTHRLKDGILFGMVGAYDMGSLVLMLEGHRLPPRMALDEBF 420
Qy      421 PPALONHAAYIGYSVSSMLRGRLFLSGAPFRHGRVIAFOUKDGAIVAAOSLOGE 480
Db      421 PPALONHAAYIGYSVSSMLRGRLFLSGAPFRHGRVIAFOUKDGAIVAAOSLOGE 480
Qy      481 QIGSYFSEBELCLDTRDGGTTDVLIVAAPMFLGPONKKTGRVYVVLVGOOSILTLQSTLQ 540
Db      481 QIGSYFSEBELCLDTRDGGTTDVLIVAAPMFLGPONKKTGRVYVVLVGOOSILTLQSTLQ 540
Qy      541 PEPPDADFPGFMAGLPDLNODGFADVAVGAPLEDHOGALYLYHGTSQVAPHPAOKIA 600
Db      541 PEPPDADFPGFMAGLPDLNODGFADVAVGAPLEDHOGALYLYHGTSQVAPHPAOKIA 600
Qy      601 AASMHAFSLYFGRSVDGRDLDDGDLVDVAVGAOGAAILLSRPIVHLTPSLEVTPOAIS 660
Db      601 AASMHAFSLYFGRSVDGRDLDDGDLVDVAVGAOGAAILLSRPIVHLTPSLEVTPOAIS 660
Qy      661 VVORDCRRRGQFAVCLTAALCFQVTSRTPGRWDHGFYMRFTASLDEMTAGARAAPDGSQ 720
Db      661 VVORDCRRRGQFAVCLTAALCFQVTSRTPGRWDHGFYMRFTASLDEMTAGARAAPDGSQ 720
Qy      721 RLSPRLSLVGNWTCQQLHFHVLDTSYLRLPVALTVPFALDNTTKRGPVLANEGSPISQ 780
Db      721 RLSPRLSLVGNWTCQQLHFHVLDTSYLRLPVALTVPFALDNTTKRGPVLANEGSPISQ 780
Qy      781 KLVPSSKDCGPNCEVTDLVQNMNDIRGSRKAPFVVGGRKVLVSTLLENKKNAYNT 840
Db      781 KLVPSSKDCGPNCEVTDLVQNMNDIRGSRKAPFVVGGRKVLVSTLLENKKNAYNT 840
Qy      841 SLSTIIFSRNLHLASITPQRESPIKYECAAPSAHAALCSVGHVFPOTGAKVTFLLEFESC 900
Db      841 SLSTIIFSRNLHLASITPQRESPIKYECAAPSAHAALCSVGHVFPOTGAKVTFLLEFESC 900
Qy      901 SSLLSQVQKGLTASSDSLERNGLTOENTASAYIOYEPHLLFSSSESTLHREVPYGT 960
Db      901 SSLLSQVQKGLTASSDSLERNGLTOENTASAYIOYEPHLLFSSSESTLHREVPYGT 960
Qy      961 FVGPGEPEFTTLR 973
Db      961 FVGPGEPEFTTLR 973

```

RESULT 4  
AAB64658 standard; Protein; 1152 AA.  
ID AAB64658  
XX  
AC AAB64658;  
XX

```

Dt      22-MAR-2001 (first entry)
XX      Human secreted protein BLAST search protein SEQ ID NO: 168.
DE      Cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;
XX      antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX      cardiovascular; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX      valiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX      neurological disease; infection; human; secreted protein.
XX
OS      Homo sapiens.
XX
PN      W0200077197-A1.
XX
PD      21-DEC-2000.
XX
PF      01-JUN-2000; 2000WO-US14934.
XX
PR      11-JUN-1999; 99US-0138599.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX      (ROSE/) ROSEN C A.
PI      Rosen CA, Ruben SM, Komatsoulis GA;
XX      WPI; 2001-032312/04.
DR
PT      Isolated nucleic acid molecule encoding a human secreted protein is
XX      used in preventing, treating or ameliorating a medical condition -
XX
PS      Disclosure: Page 547-551; 558pp; English.
XX
CC      The invention relates to the isolation of genes AAF3757-F37803 encoding
XX      the human secreted proteins AAB64549-B64594. The sequence is used as a
XX      query sequence for doing BLAST searches to identify homologous
XX      sequences. The genes and proteins are useful for preventing,
XX      ameliorating or treating medical conditions, e.g. by protein or gene
XX      therapy. The genes are isolated from a range of human tissues disclosed
XX      in the specification. The nucleic acids, proteins, antibodies and
XX      (ant) agonists are useful in the diagnosis, treatment and prevention of:
XX      (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
XX      adrenal gland, bone, bone marrow, breast, gastrointestinal
XX      tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
XX      disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX      thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
XX      rheumatoid arthritis and ulcerative colitis; (c) cardiovascular
XX      disorders such as myocardial ischaemia; (d) wound healing; (e)
XX      neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX      infectious diseases such as viral, bacterial, fungal and parasitic
XX      infections.
XX
SQ      Sequence 1152 AA;

```

Query Match 36.6%; Score 427; DB 22; Length 1152;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      2 ELFPVTHLFLPLVLTGLGCSFNLDEHHRLFPGPPEAREGYSLVQHVGGQRMMLVGA 61
Db      2 ELFPVTHLFLPLVLTGLGCSFNLDEHHRLFPGPPEAREGYSLVQHVGGQRMMLVGA 61
Qy      62 WDGSQDRGDVYRCVPGAHNAPCAKGLDYOJGNSSHPAVMHLMGSLLETDDGCG 121
Db      62 WDGSQDRGDVYRCVPGAHNAPCAKGLDYOJGNSSHPAVMHLMGSLLETDDGCG 121
Qy      122 MACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVYIVLDGNSIYP 181
Db      122 MACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVYIVLDGNSIYP 181
Qy      182 SEVOTPLRLVGLKFLIDPEQIOVGLVOYGSPVHEWSLGDPRTEEYVRAAKNLSRREG 241
Db      182 SEVOTPLRLVGLKFLIDPEQIOVGLVOYGSPVHEWSLGDPRTEEYVRAAKNLSRREG 241

```

QY 242 ETKTAQAIWVACTEGFSQSHGGRPEARLLVVTVDGSHDEBELPALKACACAGRVTRYG 301  
 DB 242 ETKTAQAIWVACTEGFSQSHGGRPEARLLVVTVDGSHDEBELPALKACACAGRVTRYG 301  
 QY 302 IAVLGHYLRORDPSSFLREIRTIASDPDERFFFNVTDEALTDIVDALGDRIFGLESGH 361  
 DB 302 IAVLGHYLRORDPSSFLREIRTIASDPDERFFFNVTDEALTDIVDALGDRIFGLESGH 361  
 QY 362 AENESSFGLMSQIGFSTRRLKDGILFGWGAAYDMGGSVLWLEGGHRLPPPRMALEDEFP 421  
 DB 362 AENESSFGLMSQIGFSTRRLKDGILFGWGAAYDMGGSVLWLEGGHRLPPPRMALEDEFP 421  
 QY 422 PALONHA 428  
 DB 422 PALONHA 428

RESULT 5  
 AAB64584  
 ID AAB64584 standard; Protein; 1167 AA.  
 XX  
 AC AAB64584;  
 XX  
 DT 22-MAR-2001 (first entry)  
 XX  
 DE Human secreted protein #37.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW caridiart; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200077197-A1.  
 XX  
 PD 21-DEC-2000.  
 XX  
 PF 01-JUN-2000; 2000WO-US14934.  
 XX  
 PR 11-JUN-1999; 99US-0138599.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis GA;  
 XX  
 DR WPI; 2001-032312/04.  
 DR N-PSDB; AAF32793.  
 PT Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 XX  
 XX Claim 11; Page 496-500; 558bp; English.  
 XX  
 PS Sequences AAB64549-B64594 represent the amino acid sequences of 47  
 CC human secreted proteins encoded by the genes AAF32757-F32803. The genes  
 CC and proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
 CC a range of human tissues disclosed in the specification. The nucleic  
 CC acids, proteins, antibodies and (ant)agonists are useful in the  
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC  
 XX

Sequence 1167 AA;  
 Query Match 36.6%; Score 427; DB 22; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ELFPVTHFLFLVPLTGLCSFPNLDENHRLFPGPPEAFEGYSVLQHVGGGQRMWLVGAP 61  
 DB 2 ELFPVTHFLFLVPLTGLCSFPNLDENHRLFPGPPEAFEGYSVLQHVGGGQRMWLVGAP 61  
 QY 62 WDPSGDRRGVYRCPCVGGANAPCAKGLDYLQNSHPAVNMHLGMSLLETDDGCGF 121  
 DB 62 WDPSGDRRGVYRCPCVGGANAPCAKGLDYLQNSHPAVNMHLGMSLLETDDGCGF 121  
 QY 122 MACAPLMSRACGSSVFSSGICARVDASFQPGSLAFTQRCPTTMDVIVLDGNSITPW 181  
 DB 122 MACAPLMSRACGSSVFSSGICARVDASFQPGSLAFTQRCPTTMDVIVLDGNSITPW 181  
 QY 182 SEVQTFRLRLVGLFIDPEQIQVGLVOYGESPVHWSLGDRTKEVVRAKNLSRRBGR 241  
 DB 182 SEVQTFRLRLVGLFIDPEQIQVGLVOYGESPVHWSLGDRTKEVVRAKNLSRRBGR 241  
 QY 242 ETKTAQAIWVACTEGFSQSHGGRPEARLLVVTVDGSHDEBELPALKACACAGRVTRYG 301  
 DB 242 ETKTAQAIWVACTEGFSQSHGGRPEARLLVVTVDGSHDEBELPALKACACAGRVTRYG 301  
 QY 302 IAVLGHYLRORDPSSFLREIRTIASDPDERFFFNVTDEALTDIVDALGDRIFGLESGH 361  
 DB 302 IAVLGHYLRORDPSSFLREIRTIASDPDERFFFNVTDEALTDIVDALGDRIFGLESGH 361  
 QY 362 AENESSFGLMSQIGFSTRRLKDGILFGWGAAYDMGGSVLWLEGGHRLPPPRMALEDEFP 421  
 DB 362 AENESSFGLMSQIGFSTRRLKDGILFGWGAAYDMGGSVLWLEGGHRLPPPRMALEDEFP 421  
 QY 422 PALONHA 428  
 DB 422 PALONHA 428

RESULT 6  
 AAU76853  
 ID AAU76853 standard; Protein; 195 AA.  
 XX  
 AC AAU76853;  
 XX  
 DT 21-MAY-2002 (first entry)  
 XX  
 DE Human integrin alpha subunit Alpha 10 A domain.  
 XX  
 KW Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;  
 KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
 KW ischaemia-reperfusion injury; immune complex; parasitic disease;  
 KW antinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200209737-A1.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 31-JUL-2001; 2001WO-US23957.  
 XX  
 PR 31-JUL-2000; 2000US-221950P.  
 PR 11-JAN-2001; 2001US-0758493.  
 PR 13-MAR-2001; 2001US-0805354.  
 XX  
 PA (GHEO ) GEN HOSPITAL CORP.  
 XX  
 PI Arnaout AM, Li R, Xiong J;  
 XX  
 DR WPI; 2002-188687/24.  
 XX  
 PT Novel high affinity integrin polypeptide useful for treating restenosis  
 XX



PT and parasitic diseases, comprises all or part of variant integrin alpha subunit A domain or variant integrin beta subunit A-like domain -

Example 2; Fig 5; 55pp; English.

CC The invention relates to a high affinity integrin polypeptide comprising  
CC all or part of a variant integrin alpha subunit A domain or a variant  
CC integrin beta subunit A-like domain. The polypeptide, preferably the  
CC CD11b alpha subunit A domain, where I at residue 332 has been replaced by  
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
CC for determining if a test compound is a candidate compound for binding to  
CC CD11b or for treating an inflammatory disorder, by contacting a test  
CC compound with the polypeptide and determining if the test compound binds  
CC to the polypeptide. The integrin subunits are useful for reducing  
CC skeletal muscle injury, for treating disorders caused by  
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
CC diseases, to purify variant integrin polypeptide ligands and as bait  
CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
CC the human integrin alpha subunit Alpha 10 A domain.

XX Sequence 195 AA;

Query Match 16.7%; Score 195; DB 23; Length 195;  
Best Local Similarity 100.0%; Pred. No. 6.3e-191;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 CPTVMDVVIYVLDGNSIYPMSEVQFLRLVGLKFLIDPEQIQVGLVYGESPVHEWSIGD 221  
DB 1 CPTVMDVVIYVLDGNSIYPMSEVQFLRLVGLKFLIDPEQIQVGLVYGESPVHEWSIGD 60  
QY 222 FRTKEEVVRAAKNLSRREGRETQTAQIMVACTEGFSQSHGREGRAARLLVVTDSGSHD 281  
DB 61 FRTKEEVVRAAKNLSRREGRETQTAQIMVACTEGFSQSHGREGRAARLLVVTDSGSHD 120  
QY 282 GEELPALKACEAGRTVRYGIAVLGHYLRORDPSFLREIRTIASDPDERFFNVTDDEA 341  
DB 121 GEELPALKACEAGRTVRYGIAVLGHYLRORDPSFLREIRTIASDPDERFFNVTDDEA 180  
QY 342 ALTDIVDALGDRIFG 356  
DB 181 ALTDIVDALGDRIFG 195

RESULT 7

AAU76862 standard; Protein; 195 AA.

XX AAU76862;

DT 21-MAY-2002 (first entry)

XX Human integrin alpha subunit Alpha 10 variant A domain.

XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;  
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;  
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;  
KW mutein.

XX Homo sapiens.  
OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 193 /note= "Wild-type Ile substituted by any other amino  
FT acid"

XX WO200209737-A1.

XX 07-FEB-2002.

XX 31-JUL-2001; 2001WO-US23957.

XX 31-JUL-2000; 2000US-221950P.  
PR 11-JAN-2001; 2001US-0758493.  
PR 13-MAR-2001; 2001US-0805354.

XX (GENO ) GEN HOSPITAL CORP.

PI Arnout AM, Li R, Xiong J;

XX WPI, 2002-188687/24.

XX Novel high affinity integrin polypeptide useful for treating restenosis  
PT and parasitic diseases, comprises all or part of variant integrin alpha  
PR subunit A domain or variant integrin beta subunit A-like domain -

XX Claim 53; Page -: 55pp; English.

XX The invention relates to a high affinity integrin polypeptide comprising  
CC all or part of a variant integrin alpha subunit A domain or a variant  
CC integrin beta subunit A-like domain. The polypeptide, preferably the  
CC CD11b alpha subunit A domain, where I at residue 332 has been replaced by  
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
CC for determining if a test compound is a candidate compound for binding to  
CC CD11b or for treating an inflammatory disorder, by contacting a test  
CC compound with the polypeptide and determining if the test compound binds  
CC to the polypeptide. The integrin subunits are useful for reducing  
CC skeletal muscle injury, for treating disorders caused by  
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
CC diseases, to purify variant integrin polypeptide ligands and as bait  
CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
CC a human integrin alpha subunit Alpha 10 variant A domain.  
CC Note: This variant sequence is not featured in the specification but has  
CC been derived from the wild-type protein shown in AAU76863.

XX Sequence 195 AA;

Query Match 16.5%; Score 192; DB 23; Length 195;  
Best Local Similarity 100.0%; Pred. No. 5.2e-178;  
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 CPTVMDVVIYVLDGNSIYPMSEVQFLRLVGLKFLIDPEQIQVGLVYGESPVHEWSIGD 221  
DB 1 CPTVMDVVIYVLDGNSIYPMSEVQFLRLVGLKFLIDPEQIQVGLVYGESPVHEWSIGD 60  
QY 222 FRTKEEVVRAAKNLSRREGRETQTAQIMVACTEGFSQSHGREGRAARLLVVTDSGSHD 281  
DB 61 FRTKEEVVRAAKNLSRREGRETQTAQIMVACTEGFSQSHGREGRAARLLVVTDSGSHD 120  
QY 282 GEELPALKACEAGRTVRYGIAVLGHYLRORDPSFLREIRTIASDPDERFFNVTDDEA 341  
DB 121 GEELPALKACEAGRTVRYGIAVLGHYLRORDPSFLREIRTIASDPDERFFNVTDDEA 180  
QY 342 ALTDIVDALGDR 353  
DB 181 ALTDIVDALGDR 192

RESULT 8

AAAY41735 standard; Protein; 124 AA.

XX AAAY41735;

DT 07-DEC-1999 (first entry)

XX Human PRO827 protein sequence.

XX Human; PRO: EST: expressed sequence tag; PCR primer; hybridisation;  
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
KW secreted protein; transmembrane protein.

XX Homo sapiens.

```

XX WO9946281-A2.
XX
XX 16-SEP-1999.
XX
XX 08-MAR-1999; 99MO-US05028.
XX
XX 10-MAR-1999; 98US-0077450.
XX 11-MAR-1999; 98US-0077632.
XX 11-MAR-1999; 98US-0077641.
XX 11-MAR-1999; 98US-0077649.
XX 12-MAR-1999; 98US-0077791.
XX 13-MAR-1999; 98US-0078004.
XX 17-MAR-1999; 98US-0040220.
XX 20-MAR-1999; 98US-0078886.
XX 20-MAR-1999; 98US-0079910.
XX 20-MAR-1999; 98US-0078936.
XX 20-MAR-1999; 98US-0078939.
XX 25-MAR-1999; 98US-0079294.
XX 26-MAR-1999; 98US-0079656.
XX 27-MAR-1999; 98US-0079663.
XX 27-MAR-1999; 98US-0079664.
XX 27-MAR-1999; 98US-0079689.
XX 27-MAR-1999; 98US-0079728.
XX 27-MAR-1999; 98US-0079786.
XX 30-MAR-1999; 98US-0079920.
XX 30-MAR-1999; 98US-0079923.
XX 31-MAR-1999; 98US-0080105.
XX 31-MAR-1999; 98US-0080107.
XX 31-MAR-1999; 98US-0080165.
XX 31-MAR-1999; 98US-0080194.
XX 01-APR-1999; 98US-0080327.
XX 01-APR-1999; 98US-0080328.
XX 01-APR-1999; 98US-0080333.
XX 01-APR-1999; 98US-0080334.
XX 08-APR-1999; 98US-0081049.
XX 08-APR-1999; 98US-0081070.
XX 08-APR-1999; 98US-0081071.
XX 09-APR-1999; 98US-0081195.
XX 09-APR-1999; 98US-0081203.
XX 09-APR-1999; 98US-0081229.
XX 15-APR-1999; 98US-0081817.
XX 15-APR-1999; 98US-0081838.
XX 15-APR-1999; 98US-0081952.
XX 15-APR-1999; 98US-0081955.
XX 21-APR-1999; 98US-0082568.
XX 21-APR-1999; 98US-0082569.
XX 22-APR-1999; 98US-0082700.
XX 22-APR-1999; 98US-0082704.
XX 22-APR-1999; 98US-0082804.
XX 23-APR-1999; 98US-0082767.
XX 23-APR-1999; 98US-0082796.
XX 27-APR-1999; 98US-0083336.
XX 28-APR-1999; 98US-0083322.
XX 29-APR-1999; 98US-0083392.
XX 29-APR-1999; 98US-0083495.
XX 29-APR-1999; 98US-0083496.
XX 29-APR-1999; 98US-0083499.
XX 29-APR-1999; 98US-0083500.
XX 29-APR-1999; 98US-0083545.
XX 29-APR-1999; 98US-0083554.
XX 29-APR-1999; 98US-0083558.
XX 29-APR-1999; 98US-0083559.
XX 30-APR-1999; 98US-0083742.
XX 05-MAY-1999; 98US-0084366.
XX 06-MAY-1999; 98US-0084414.
XX 07-MAY-1999; 98US-0084441.
XX 07-MAY-1999; 98US-0084598.
XX 07-MAY-1999; 98US-0084600.
XX 07-MAY-1999; 98US-0084627.
XX 07-MAY-1999; 98US-0084637.
XX 07-MAY-1999; 98US-0084639.
XX 07-MAY-1999; 98US-0084640.

```

---

```

PR 07-MAY-1999; 98US-0084643.
PR 13-MAY-1999; 98US-0085323.
PR 13-MAY-1999; 98US-0085338.
PR 13-MAY-1999; 98US-0085339.
PR 15-MAY-1999; 98US-0085379.
PR 15-MAY-1999; 98US-0085580.
PR 15-MAY-1999; 98US-0085582.
PR 15-MAY-1999; 98US-0085689.
PR 15-MAY-1999; 98US-0085697.
PR 15-MAY-1999; 98US-0085700.
PR 15-MAY-1999; 98US-0085704.
PR 18-MAY-1999; 98US-0086023.
PR 22-MAY-1999; 98US-0086392.
PR 22-MAY-1999; 98US-0086414.
PR 22-MAY-1999; 98US-0086430.
PR 22-MAY-1999; 98US-0086486.
PR 28-MAY-1999; 98US-0087099.
PR 28-MAY-1999; 98US-0087106.
PR 28-MAY-1999; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
XX (GETH ) GENENTECH INC.
XX
XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
XX WPI: 1999-551358/46.
XX N-PSDB; AAZ34185.
XX
XX New secreted and transmembrane polypeptides and their polynucleotides,
XX useful for treating blood coagulation disorders, cancers and cellular
XX adhesion disorders -
XX
XX Claim 12, Fig 139; 530pp; English.
XX
XX The present invention describes secreted and transmembrane polypeptides
XX and their polynucleotides. The nucleotide sequences are useful as
XX sources of probes, primers, for chromosome mapping, and for generation
XX of antisense sequences. They can also be used to create transgenic
XX animals. The proteins can be used to treat a variety of diseases and
XX disorders, depending on their function. Diseases that may be treated
XX include blood coagulation disorders, cancers and cellular adhesion
XX disorders. They may also be used to raise antibodies. AAZ33891 to
XX AAZ43338, and AA41685 to AA41774 represent polynucleotide and
XX polypeptide sequence given in the exemplification of the present
XX invention.
XX
XX Sequence 124 AA:
XX
SQ

```

Query Match 10.5%; Score 122; DB 20; Length 124;  
Best Local Similarity 100.0%; Pred. No. 46e-110; Mismatches 0; Indels 0; Gaps 0;  
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MELPFVTHLFLPLVFLTGLGSPFNLDHHPRLFRGPEAEFGYSVLQHVGGGQRMVLVGA 60
DB 1 MELPFVTHLFLPLVFLTGLGSPFNLDHHPRLFRGPEAEFGYSVLQHVGGGQRMVLVGA 60
QY 61 PWDGSGDRGDDVYRCVPVGAHNAKAGHIGDYQLGNSSHPAVNMHLGMSLLETDGCG 120
DB 61 PWDGSGDRGDDVYRCVPVGAHNAKAGHIGDYQLGNSSHPAVNMHLGMSLLETDGCG 120
QY 121 FM 122
DB 121 FM 122

```

RESULT 9  
AAB44291  
ID AAB44291 standard; Protein: 124 AA.  
AC AAB44291;  
XX

```

DT 08-FEB-2001 (first entry)
XX
XX Human PRO827 (UNQ468) protein sequence SEQ ID NO:346.
XX
XX Human; secreted protein, transmembrane protein; PRO; EST; cytosolic;
XX expressed sequence tag; detection; cancer.
XX
XX Homo sapiens.
XX
XX WO200053756-A2.
XX
XX 14-SEP-2000.
XX
XX 18-FEB-2000; 2000WO-US04341.
XX
XX 08-MAR-1999; 99WO-US05028.
XX 12-MAR-1999; 99US-0123957.
XX 29-MAR-1999; 99US-0126773.
XX 21-APR-1999; 99US-0130232.
XX 28-APR-1999; 99US-0131445.
XX 14-MAY-1999; 99US-0134287.
XX 23-JUN-1999; 99US-0141037.
XX 26-JUL-1999; 99US-0145698.
XX 29-OCT-1999; 99US-0162506.
XX 30-NOV-1999; 99WO-US28313.
XX 02-DEC-1999; 99WO-US28565.
XX 16-DEC-1999; 99WO-US30095.
XX 30-DEC-1999; 99WO-US31243.
XX 30-DEC-1999; 99WO-US31274.
XX 05-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00277.
XX 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnovers L, Eaton DL,
XX Ferraz N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME,
XX Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ,
XX Kijavini IJ, Kuo SS, Napier MA, Pan J, Peoni NF, Roy MA,
XX Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX WPI; 2000-611443/58.
XX N-PSDB; AAC78545.
XX
XX Novel PRO polypeptides and polynucleotides used in detection methods,
XX to target bioactive molecules to specific cells, and to modulate
XX cellular activities -
XX
XX Claim 12; Fig 139; 636pp; English.
XX
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
XX sequence tag) sequences which encode secreted or transmembrane PRO
XX polypeptides. The PRO polynucleotides and polypeptides have cytostatic
XX activity. The polynucleotides and polypeptides can be used for detecting
XX the presence of PRO polypeptides in samples, for linking bioactive
XX molecules to cells and for modulating biological activities of cells,
XX using the polypeptides for specific targeting. The polypeptide targeting
XX can be used to kill the target cells, e.g. for the treatment of cancers.
XX The polypeptide pairs provide specific targeting of bioactive molecules
XX to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
XX the isolation of the PRO polynucleotide sequences.
XX
XX Sequence 124 AA;
XX
XX Query Match 10.5%; Score 122; DB 21; Length 124;
XX Best Local Similarity 100.0%; Pred. No. 4.6e-110;
XX Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MELPVTLLPLPLVLTGLGSPFNLDHHPRLPGPPRAERQSYLVHVGGRWLVGA 60
XX 1 MELPVTLLPLPLVLTGLGSPFNLDHHPRLPGPPRAERQSYLVHVGGRWLVGA 60

```

```

Qy 61 FWDGPGSDRRGDVYRCVGGAHNAPCAKGHIGDYQLONSSHPAVNMLGMSLLETDDGCG 120
Db 61 FWDGPGSDRRGDVYRCVGGAHNAPCAKGHIGDYQLONSSHPAVNMLGMSLLETDDGCG 120
Qy 121 FM 122
Db 121 FM 122

RESULT 10
AAU29100
ID AAU29100 standard; Protein; 124 AA.
XX
XX AAU29100;
XX
XX 18-DEC-2001 (first entry)
XX
XX Human PRO polypeptide sequence #77.
XX
XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
XX dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
XX blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
XX adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200168848-A2.
XX
XX 20-SEP-2001.
XX
XX 28-FEB-2001; 2001WO-US06520.
XX
XX 01-MAR-2000; 2000WO-US05601.
XX 02-MAR-2000; 2000WO-US05841.
XX 03-MAR-2000; 2000US-187202P.
XX 06-MAR-2000; 2000US-186968P.
XX 14-MAR-2000; 2000US-189320P.
XX 14-MAR-2000; 2000US-189328P.
XX 15-MAR-2000; 2000WO-US06884.
XX 21-MAR-2000; 2000US-190828P.
XX 21-MAR-2000; 2000US-191007P.
XX 21-MAR-2000; 2000US-191048P.
XX 21-MAR-2000; 2000US-191314P.
XX 28-MAR-2000; 2000US-192655P.
XX 29-MAR-2000; 2000US-193032P.
XX 29-MAR-2000; 2000US-193053P.
XX 30-MAR-2000; 2000WO-US08439.
XX 04-APR-2000; 2000US-194449P.
XX 04-APR-2000; 2000US-194647P.
XX 11-APR-2000; 2000US-195975P.
XX 11-APR-2000; 2000US-196000P.
XX 11-APR-2000; 2000US-196187P.
XX 11-APR-2000; 2000US-196690P.
XX 11-APR-2000; 2000US-196820P.
XX 18-APR-2000; 2000US-198121P.
XX 18-APR-2000; 2000US-198585P.
XX 25-APR-2000; 2000US-199397P.
XX 25-APR-2000; 2000US-199550P.
XX 25-APR-2000; 2000US-199654P.
XX 03-MAY-2000; 2000US-201516P.
XX 17-MAY-2000; 2000WO-US13705.
XX 22-MAY-2000; 2000WO-US14042.
XX 30-MAY-2000; 2000WO-US14941.
XX 02-JUN-2000; 2000WO-US15264.
XX 05-JUN-2000; 2000US-209832P.
XX 28-JUL-2000; 2000WO-US20710.
XX 22-AUG-2000; 2000US-0644848.
XX 24-AUG-2000; 2000WO-US23328.
XX 08-NOV-2000; 2000WO-US30952.
XX 01-DEC-2000; 2000WO-US32678.
XX 20-DEC-2000; 2000WO-US34956.
XX
XX (GETH ) GENENTECH INC.

```

XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
 XX WPI: 2001-602746/68.  
 DR N-PSDB; AAS46001.  
 XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
 PT presence of tumours, such as prostate and breast tumours, in mammals and  
 PT to screen for modulators of the compounds -  
 XX Claim 11; Fig 154; 774pp; English.  
 XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.  
 CC The PRO polypeptides and their associated nucleic acids can be used to  
 CC detect the presence of a tumour in a mammal by comparing the level of  
 CC expression of a PRO polypeptide in a test sample of cells from the animal  
 CC and a control sample of normal cells, whereby a higher level of  
 CC expression in the test sample indicates the presence of a tumour in the  
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats  
 CC and rabbits but are preferably human. The polypeptides can be used to  
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,  
 CC when contacted with it. A specific polypeptide can be used to stimulate  
 CC the proliferation or differentiation of chondrocyte cells. The PRO  
 CC proteins can be used to determine the presence of tumours and also  
 CC susceptibility to tumour development, particularly adrenal, lung, colon,  
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
 CC can be used for genetic analysis of individuals with genetic disorders.  
 XX Sequence 124 AA;  
 SQ  
 Query Match 10.5%; Score 122; DB 22; Length 124;  
 Best Local Similarity 100.0%; Pred No. 4,6e-110;  
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MELPVTHTLFLPLVLTGLCSPPNLDHHPRLFPGPPEAFEGYSVLQHVGGGQRMVLVGA 60  
 DB 1 MELPVTHTLFLPLVLTGLCSPPNLDHHPRLFPGPPEAFEGYSVLQHVGGGQRMVLVGA 60  
 QY 61 PWDGSGDRGDRGVYRCPVGAGNAPCAKHLGDYQIGNSHRAVNMHLMGSLLETDGOG 120  
 DB 61 PWDGSGDRGDRGVYRCPVGAGNAPCAKHLGDYQIGNSHRAVNMHLMGSLLETDGOG 120  
 QY 121 FM 122  
 DB 121 FM 122  
 RESULT 11  
 ABB95485  
 ID ABB95485 standard; Protein; 124 AA.  
 AC ABB95485;  
 DT 19-JUL-2002 (first entry)  
 XX Human angiogenesis related protein PRO827 SEQ ID NO: 126.  
 DE Human angiogenesis related protein PRO827 SEQ ID NO: 126.  
 XX Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;  
 KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;  
 KW cardiac; cytosolic; antiangiogenic; hypotensive; vulnery;  
 KW antiarteriosclerotic.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO200208284-A2.  
 XX 31-JAN-2002.  
 PD 09-JUL-2001; 2001WO-US21735.  
 PF 20-JUL-2000; 2000US-219556P.  
 PR

PR 25-JUL-2000; 2000US-220624P.  
 PR 25-JUL-2000; 2000US-220644P.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 02-AUG-2000; 2000US-222695P.  
 PR 17-AUG-2000; 2000US-0643657.  
 PR 23-AUG-2000; 2000WO-US23522.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 07-SEP-2000; 2000US-230978P.  
 PR 15-SEP-2000; 2000US-000000P.  
 PR 18-SEP-2000; 2000US-0664610.  
 PR 18-SEP-2000; 2000US-0665350.  
 PR 24-OCT-2000; 2000US-242922P.  
 PR 08-NOV-2000; 2000US-0709238.  
 PR 10-NOV-2000; 2000WO-US30952.  
 PR 01-DEC-2000; 2000WO-US30873.  
 PR 20-DEC-2000; 2000US-0747259.  
 PR 20-DEC-2000; 2000WO-US34956.  
 PR 22-JAN-2001; 2001US-0767609.  
 PR 28-FEB-2001; 2001US-0796498.  
 PR 28-FEB-2001; 2001WO-US06520.  
 PR 01-MAR-2001; 2001WO-US06656.  
 PR 09-MAR-2001; 2001US-0802706.  
 PR 14-MAR-2001; 2001US-0808689.  
 PR 22-MAR-2001; 2001US-0816744.  
 PR 05-APR-2001; 2001US-0828366.  
 PR 10-MAY-2001; 2001US-0854208.  
 PR 10-MAY-2001; 2001US-0854280.  
 PR 25-MAY-2001; 2001US-0866028.  
 PR 25-MAY-2001; 2001US-0866034.  
 PR 25-MAY-2001; 2001WO-US17092.  
 PR 30-MAY-2001; 2001US-0870574.  
 PR 30-MAY-2001; 2001WO-US17443.  
 PR 01-JUN-2001; 2001WO-US17800.  
 PR 20-JUN-2001; 2001WO-US19692.  
 PR 28-JUN-2001; 2001WO-US00000.  
 XX (GETH) GENENTECH INC.  
 PA (BAKE) BAKER K P.  
 PA (FERR) FERRARA N.  
 PA (GERB) GERBER H.  
 PA (GERR) GERRITSEN M E.  
 PA (GODD) GODDARD A.  
 PA (GODO) GODOWSKI P J.  
 PA (GURN) GURNEY A L.  
 PA (HILL) HILLAN K J.  
 PA (MARS) MARSTERS S A.  
 PA (PANU) PAN J.  
 PA (PAON) PAONI N F.  
 PA (STEP) STEPHAN J F.  
 PA (WATA) WATANABE C K.  
 PA (WILL) WILLIAMS P M.  
 PA (WOOD) WOOD W I.  
 XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A,  
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF,  
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;  
 XX WPI: 2002-171999/22.  
 DR N-PSDB; ABL95623.  
 XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,  
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial  
 PT infarction), endothelial or angiogenic disorders in a mammal -  
 XX Claim 11; Fig 126; 567pp; English.  
 XX The present invention provides the protein and coding sequences of human  
 CC PRO proteins. These are useful for treating or diagnosing a  
 CC cardiovascular, endothelial or angiogenic disorder, including cardiac  
 CC hypertrophy, trauma, cancer, age-related macular degeneration,  
 CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,  
 CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour

CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound  
 CC healing. The present sequence is a PRO protein of the invention.  
 XX

SQ Sequence 124 AA;

Query Match 10.5%; Score 122; DB 23; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-110;  
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLFPGPEAEFGYSVLQHVGGGQRMVLVGA 60  
 DB 1 MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLFPGPEAEFGYSVLQHVGGGQRMVLVGA 60  
 OY 61 PWDGSGDRGDVYRCVPYGAAHNAAPCAKGLGDYQLGNSHPAVNMHLGMSLLETGDCGG 120  
 DB 61 PWDGSGDRGDVYRCVPYGAAHNAAPCAKGLGDYQLGNSHPAVNMHLGMSLLETGDCGG 120  
 OY 121 FM 122  
 DB 121 FM 122

RESULT 12  
 ABB84879 ID ABB84879 standard; Protein; 124 AA.

AC ABB84879;

XX 16-MAY-2002 (first entry)

DE Human PRO827 protein sequence SEQ ID NO:126.

KW Human; angiogenesis; cardiac; cytoskeletal; antiangiogenic; hypotensive;  
 KW vulnerability; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;  
 KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;  
 KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;  
 KW age-related macular degeneration; arterial restenosis; angina;  
 KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;  
 KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;  
 KW wound healing; chromosome mapping; gene mapping.

OS Homo sapiens.

XX WO200200630-A2.

XX 03-JAN-2002.

XX 20-JUN-2001; 2001WO-US19692.

XX 23-JUN-2000; 2000US-213637P.  
 PR 20-JUL-2000; 2000US-219556P.  
 PR 25-JUL-2000; 2000US-220624P.  
 PR 25-JUL-2000; 2000US-220664P.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 02-AUG-2000; 2000US-222659P.  
 PR 17-AUG-2000; 2000US-0643657.  
 PR 23-AUG-2000; 2000WO-US23322.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 07-SEP-2000; 2000US-230978P.  
 PR 18-SEP-2000; 2000US-0664610.  
 PR 18-SEP-2000; 2000US-0665350.  
 PR 24-OCT-2000; 2000US-242922P.  
 PR 08-NOV-2000; 2000US-0709238.  
 PR 08-NOV-2000; 2000WO-US10952.  
 PR 10-NOV-2000; 2000WO-US10873.  
 PR 01-DEC-2000; 2000WO-US32678.  
 PR 20-DEC-2000; 2000US-0747259.  
 PR 22-JAN-2001; 2000WO-US34956.  
 PR 28-FEB-2001; 2001US-0767609.  
 PR 28-FEB-2001; 2001WO-US06520.  
 PR 01-MAR-2001; 2001WO-US06666.  
 PR 09-MAR-2001; 2001US-0802706.

PR 14-MAR-2001; 2001US-0808689.  
 PR 22-MAR-2001; 2001US-0816744.  
 PR 05-APR-2001; 2001US-0828366.  
 PR 10-MAY-2001; 2001US-0854208.  
 PR 10-MAY-2001; 2001US-0854280.  
 PR 25-MAY-2001; 2001US-0866029.  
 PR 25-MAY-2001; 2001US-0866034.  
 PR 25-MAY-2001; 2001WO-US17092.  
 PR 30-MAY-2001; 2001US-0870574.  
 PR 30-MAY-2001; 2001WO-US17443.  
 PR 01-JUN-2001; 2001WO-US17800.

XX (GENTH ) GENENTECH INC.

PI Baker KP, Ferrara N, Gerber H, Gertsen ME, Goddard A,  
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF,  
 PI Stephan UF, Watanabe CK, Williams PM, Wood WI, Ye W;

XX WPI; 2002-090516/12.

DR N-PSDB; ABL88134.

XX One hundred and eighty seven nucleic acids encoding PRO polypeptides;  
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial  
 PT infarction), endothelial or angiogenic disorders in a mammal -

PS Claim 11; Fig 126; 565pp; English.

CC ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to  
 CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytoskeletal,  
 CC antiangiogenic, hypotensive, vulnerary and antiarteriosclerotic  
 CC activities, and can be used in gene therapy. The PRO polynucleotides,  
 CC proteins, agonists and antagonists are useful for treating or diagnosing  
 CC a cardiovascular, endothelial or angiogenic disorder in a mammal,  
 CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular  
 CC degeneration, atherosclerosis, hypertension, arterial restenosis,  
 CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,  
 CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver  
 CC carcinoma) and wound healing. The PRO polynucleotides have applications  
 CC in molecular biology, including use as hybridisation probes, and in  
 CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and  
 CC probes used in the exemplification of the present invention.

XX SQ Sequence 124 AA;

Query Match 10.5%; Score 122; DB 23; Length 124;  
 Best Local Similarity 100.0%; Pred. NO. 4.6e-110;  
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLFPGPEAEFGYSVLQHVGGGQRMVLVGA 60  
 DB 1 MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLFPGPEAEFGYSVLQHVGGGQRMVLVGA 60

OY 61 PWDGSGDRGDVYRCVPYGAAHNAAPCAKGLGDYQLGNSHPAVNMHLGMSLLETGDCGG 120  
 DB 61 PWDGSGDRGDVYRCVPYGAAHNAAPCAKGLGDYQLGNSHPAVNMHLGMSLLETGDCGG 120

OY 121 FM 122  
 DB 121 FM 122

RESULT 13  
 AAY32282 ID AAY32282 standard; Protein; 303 AA.

XX AAY32282;

XX 15-FEB-2000 (first entry)

DE Mouse integrin subunit alpha-10 partial sequence.

XX Integrin alpha-10; ISa10; mouse; trauma; rheumatoid arthritis;  
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;

KW inflammation; therapy; cartilage; chondrocyte; osteoblast;  
 XX fibroblast; vaccine; marker.  
 OS Mus musculus.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 15  
 FT /note= "encoded by TWC"  
 XX  
 XX WO9951639-A1.  
 XX  
 XX 14-OCT-1999.  
 XX  
 XX 31-MAR-1999; 99WO-SE00544.  
 XX  
 XX 02-APR-1998; 98SE-0001164.  
 XX 28-JAN-1999; 99SE-0000319.  
 XX  
 XX (ACT1-) ACTIVE BIOTECH AB.  
 XX  
 XX Lundgren-Akerlund E;  
 XX  
 XX WPI, 2000-052639/04.  
 XX N-PSDB; AA234835.  
 XX  
 XX New isolated integrin subunit alpha-10, used as a marker or target  
 PT molecule for cells during development, regeneration and pathological  
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 PT inflammation -  
 XX  
 XX Example 15; Fig 15; 90pp; English.  
 XX  
 XX This sequence represents a polypeptide encoded by an isolated  
 CC partial genomic DNA (see AA234835) of murine integrin subunit  
 CC alpha-10 (ISa10). The mouse genomic alpha-10 sequence was used  
 CC to generate a targeting vector for knockout experiments. The  
 CC invention relates to a recombinant or isolated integrin heterodimer  
 CC comprising the alpha10 subunit in association with subunit beta  
 CC (especially beta-1). The heterodimer and the subunit alpha-10 can  
 CC be used as markers or targets of all types of cells, e.g. of  
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used  
 CC for treating pathological conditions involving ISa10, such as:  
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
 CC for detecting the formation of cartilage during embryonal  
 CC development, physiological or therapeutic repair of cartilage,  
 CC or detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes; for selection and  
 CC analysis or for sorting, isolating or purification of chondrocytes  
 CC and for in vitro studies of differentiation of chondrocytes; and as  
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
 CC skeletal muscle or other tissues where adhesion impairs the function  
 CC of the tissue (all claimed). Binding entities, polynucleotides  
 CC (PNE) or oligonucleotides (ONS) can be used for e.g. determining the  
 CC differentiation-state of cells during embryonic development,  
 CC angiogenesis or development of cancer, in pathological conditions  
 CC such as rheumatoid arthritis, osteoarthritis or cancer, in tissue  
 CC regeneration or in therapeutic and physiological repair of  
 CC cartilage (claimed). The PNEs and ONS can also be used for detecting  
 CC pathological conditions such as atherosclerosis or inflammation  
 CC (claimed). A vaccine comprising the integrin heterodimer or  
 CC subunit alpha-10 is also claimed. Vectors, host cells and methods  
 CC of producing recombinant ISa10 are provided.  
 CC  
 CC Sequence 303 AA;  
 CC  
 CC Query Match 3.6%; Score 42; DB 21; Length 303;  
 CC Best Local Similarity 100.0%; Pred. No. 7.2e-32;  
 CC Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC 35 GPPEAFGYSVLOHVGCGORMLVGAPMDGSGDRGDYRC 76  
 CC ||||||||||||||||||||||||||||||||||||||||  
 CC 36 GPPEAFGYSVLOHVGCGORMLVGAPMDGSGDRGDYRC 77

RESULT 14  
 ID AAY32244 standard; Peptide: 22 AA.  
 XX AAY32244;  
 XX  
 XX AAY32244;  
 XX  
 XX 15-FEB-2000 (first entry)  
 XX  
 XX Human integrin subunit alpha-10 cytoplasmic domain peptide.  
 XX  
 XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;  
 KW fibroblast; vaccine; marker.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO9951639-A1.  
 XX  
 XX 14-OCT-1999.  
 XX  
 XX 31-MAR-1999; 99WO-SE00544.  
 XX  
 XX 02-APR-1998; 98SE-0001164.  
 XX 28-JAN-1999; 99SE-0000319.  
 XX  
 XX (ACT1-) ACTIVE BIOTECH AB.  
 XX  
 XX Lundgren-Akerlund E;  
 XX  
 XX WPI, 2000-052639/04.  
 XX  
 XX New isolated integrin subunit alpha-10, used as a marker or target  
 PT molecule for cells during development, regeneration and pathological  
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 PT inflammation -  
 XX  
 XX Claim 21; Page 53; 90pp; English.  
 XX  
 XX This sequence represents a fragment of novel human chondrocyte  
 CC integrin subunit alpha-10 (ISa10, see AAY32242), corresponding to  
 CC the C-terminal cytoplasmic domain of the protein. The invention  
 CC relates to a recombinant or isolated integrin heterodimer  
 CC comprising the alpha10 subunit in association with subunit beta  
 CC (especially beta-1). The integrin heterodimer, or the subunit  
 CC alpha-10, or a fragment of it such as the present sequence, can be  
 CC used as a marker or target of all types of cells, e.g. of  
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used:  
 CC for treating pathological conditions involving ISa10, such as  
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;  
 CC for detecting the formation of cartilage during embryonal  
 CC development, physiological or therapeutic repair of cartilage,  
 CC or detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes; for selection and  
 CC analysis or for sorting, isolating or purification of chondrocytes  
 CC and for in vitro studies of differentiation of chondrocytes; and as  
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,  
 CC skeletal muscle or other tissues where adhesion impairs the function  
 CC of the tissue (all claimed). ISa10 binding entities can be used to  
 CC determine the differentiation-state of cells during embryonic  
 CC development, angiogenesis or development of cancer, in pathological  
 CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,  
 CC in tissue regeneration or in therapeutic and physiological repair of  
 CC cartilage (claimed). A vaccine comprising the integrin  
 CC heterodimer or subunit alpha-10 is also claimed. ISa10  
 CC polynucleotides, vectors, host cells and methods of producing  
 CC recombinant ISa10 are also claimed.  
 CC  
 CC Sequence 22 AA;  
 CC  
 CC Query Match 1.9%; Score 22; DB 21; Length 22;  
 CC Best Local Similarity 100.0%; Pred. No. 2.1e-13;  
 CC

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1146 KLGFFAHKKIPPEEKREKLEQ 1167  
 |||||  
 DB 1 KLGFFAHKKIPPEEKREKLEQ 22

RESULT 15  
 AAY32245  
 ID AAY32245 standard; Peptide; 18 AA.  
 XX AAY32245;  
 AC AAY32245;  
 XX  
 DT 15-FEB-2000 (first entry)  
 XX  
 DE Bovine integrin subunit alpha-10 tryptic peptide 1.  
 XX  
 KM Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
 KM osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
 KM inflammation; therapy; cartilage; chondrocyte; osteoblast;  
 KM fibroblast; vaccine; marker.  
 XX  
 OS Bos taurus.  
 XX  
 PN WO9951639-A1.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 31-MAR-1999; 99WO-SE00544.  
 XX  
 PR 02-APR-1998; 98SE-0001164.  
 PR 28-JAN-1999; 99SE-0000319.  
 XX  
 PA (ACTI-) ACTIVE BIOTECH AB.  
 XX  
 PI Lundgren-Akerlund E;  
 DR WPI; 2000-052639/04.  
 XX  
 PT New isolated integrin subunit alpha-10, used as a marker or target  
 PT molecule for cells during development, regeneration and pathological  
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 PT inflammation -  
 XX  
 PS Example 2; Fig 2; 90P; English.  
 XX  
 CC This sequence represents tryptic peptide 1 of integrin subunit  
 CC alpha-10 (ISa10) isolated from bovine chondrocytes. 6 tryptic  
 CC peptides (see AAY32245-50) were obtained. PCR primers (see  
 CC AAZ34721-24) based on peptides 1 and 2 were used to obtain bovine  
 CC ISa10 cDNA fragments. These were used to screen a human  
 CC articular chondrocyte cDNA library in order to isolate human ISa10  
 CC cDNA (see AAZ34719). The invention relates to a recombinant or  
 CC isolated integrin heterodimer comprising the alpha-10 subunit in  
 CC association with subunit beta (especially beta-1). The heterodimer  
 CC or the subunit alpha-10 can be used as a marker or target of all  
 CC types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.  
 CC They can also be used as active ingredients in pharmaceutical  
 CC compositions and vaccines.  
 CC  
 XX  
 SQ Sequence 18 AA;  
 Query Match 1.2%; Score 14; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 927 NTAQTSAYIQYEPH 940  
 |||||  
 DB 2 NTAQTSAYIQYEPH 15

RESULT 16  
 AAU87675

ID AAU87675 standard; Protein; 103 AA.  
 XX  
 AC AAU87675;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Novel central nervous system protein #585.  
 XX  
 KM Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KM hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 KM cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;  
 KM nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KM acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KM adenocarcinoma; reproductive system disorder; testicular feminization;  
 KM endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 KM respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KM myocardial infarction; wound healing; cell proliferation; skin aging;  
 KM food additive; food preservative; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20015318-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01332.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0228929.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.

[illegible]

Query Match	1.1k: Score 13; DB 22; Length 103;
Best Local Similarity	100.0%; Pred. NO. 0.00047; Indels 0; Gaps 0.
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
QY 169 VIIVDGSNSIYPM 181	
DB 15 VIIVDGSNSIYPM 27	
RESULT 17	
AAU19822	
AAU19822 standard; Protein; 103 AA.	
AAU19822;	
06-DEC-2001 (first entry)	
Human novel extracellular matrix protein, Seq ID No 472.	











PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225265.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226881.  
 PR 22-AUG-2000; 2000US-0226886.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0233399.  
 PR 14-SEP-2000; 2000US-0233400.  
 PR 14-SEP-2000; 2000US-0233401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249269.  
 PR 17-NOV-2000; 2000US-0249270.  
 PR 17-NOV-2000; 2000US-0249271.  
 PR 17-NOV-2000; 2000US-0249272.  
 PR 17-NOV-2000; 2000US-0249273.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250161.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251968.  
 PR 05-DEC-2000; 2000US-0251969.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251858.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-465572/50.  
 DR N-PSDB; AMS31365.  
 XX  
 PT Nucleic acid molecules encoding human secreted extracellular matrix  
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
 PT Alzheimer's and Parkinson's diseases and cancers -  
 XX  
 PS Claim 11; SEQ ID No 444; 577bp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules encoding  
 CC novel human secreted extracellular matrix proteins (SPs). The  
 CC polynucleotides and proteins are used to prevent, treat a medical  
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
 CC chickens or sheep. For example, disorders associated with decreased  
 CC expression of SPs. The SP polynucleotide or a vector expressing them may  
 CC be administered to treat diseases by gene therapy. Antisense molecules  
 CC may be administered to down regulate expression of SPs by binding with  
 CC the cells own genes and preventing their expression. The polynucleotides  
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
 CC used as antigens to produce antibodies and to identify modulators  
 CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and  
 CC antagonists may also be used to down regulate expression and activity of  
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.

CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
 CC Caucher's disease), neurological diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease), cardio-/cerebrovascular disorders (e.g. cardiac  
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
 CC include wound healing, maintenance of organs before transplantation,

Query Match 1.1%; Score 13; DB 22; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 0.00064;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181  
 |||||  
 Db 49 VIVLDGNSIYPM 61

RESULT 21

ID ABP47854 standard; Protein; 148 AA.

XX AC ABP47854;

XX DT 23-AUG-2002 (first entry)

XX DE Human polypeptide SEQ ID NO 284.

XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antiskilling; antianaemic; antiaxitic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antinflammatory;  
 KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotoxic; gene therapy; vaccine.

XX OS Homo sapiens.

XX PN US2002042386-A1.

XX PD 11-APR-2002.

XX PF 17-JAN-2001; 2001US-0764870.

XX PR 31-JAN-2000; 2000US-179065P.

XX PR 04-FEB-2000; 2000US-180628P.

XX PR 28-JUN-2000; 2000US-214866P.

XX PR 07-JUL-2000; 2000US-216647P.

XX PR 11-JUL-2000; 2000US-216880P.

XX PR 11-JUL-2000; 2000US-217487P.

XX PR 14-JUL-2000; 2000US-218290P.

XX PR 26-JUL-2000; 2000US-220963P.

XX PR 26-JUL-2000; 2000US-220964P.

XX PR 14-AUG-2000; 2000US-224518P.

XX PR 14-AUG-2000; 2000US-224519P.

XX PR 14-AUG-2000; 2000US-225267P.

XX PR 14-AUG-2000; 2000US-225268P.

XX PR 14-AUG-2000; 2000US-225270P.

XX PR 14-AUG-2000; 2000US-225447P.

XX PR 14-AUG-2000; 2000US-225757P.

XX PR 14-AUG-2000; 2000US-225758P.

XX PR 22-AUG-2000; 2000US-226688P.

XX PR 30-AUG-2000; 2000US-228924P.

XX PR 01-SEP-2000; 2000US-229267P.

XX PR 01-SEP-2000; 2000US-229343P.

XX PR 01-SEP-2000; 2000US-229344P.

XX PR 01-SEP-2000; 2000US-229345P.

XX PR 05-SEP-2000; 2000US-229509P.

XX PR 05-SEP-2000; 2000US-229513P.

XX PR 08-SEP-2000; 2000US-231413P.

XX PR 21-SEP-2000; 2000US-234223P.

PR 21-SEP-2000; 2000US-234274P.  
 PR 25-SEP-2000; 2000US-234997P.  
 PR 27-SEP-2000; 2000US-235834P.  
 PR 29-SEP-2000; 2000US-236327P.  
 PR 29-SEP-2000; 2000US-236367P.  
 PR 29-SEP-2000; 2000US-236368P.  
 PR 29-SEP-2000; 2000US-236369P.  
 PR 29-SEP-2000; 2000US-236370P.  
 PR 02-OCT-2000; 2000US-236802P.  
 PR 02-OCT-2000; 2000US-237037P.  
 PR 02-OCT-2000; 2000US-237038P.  
 PR 02-OCT-2000; 2000US-237039P.  
 PR 02-OCT-2000; 2000US-237040P.  
 PR 13-OCT-2000; 2000US-239935P.  
 PR 20-OCT-2000; 2000US-240960P.  
 PR 20-OCT-2000; 2000US-241785P.  
 PR 20-OCT-2000; 2000US-241809P.  
 PR 01-NOV-2000; 2000US-244617P.  
 PR 17-NOV-2000; 2000US-249299P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251869P.

XX PA (ROSE/) ROSEN C A.  
 XX PA (RUBE/) RUBEN S M.  
 XX PA (BARA/) BARASH S C.

XX PI Rosen CA, Ruben SM, Barash SC;

XX DR WPI, 2002-470713/50.

XX DR N-PSDB; ABQ66529.

PT New nucleic acid encoding human proteins, useful for diagnosis,  
 PT treatment and prevention of e.g. osteoporosis, also related  
 PT polypeptides and antibodies -

PS Claim 11; SEQ ID NO 284; 215pp + Sequence listing; English.

XX The invention relates to novel genes (ABQ66521-ABQ66785) and proteins  
 CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=99909764870.

XX Sequence 148 AA;

Query Match 1.1%; Score 13; DB 23; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 0.00064;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181  
 |||||  
 Db 49 VIVLDGNSIYPM 61

RESULT 22

ID ABP48014 standard; Protein; 148 AA.

XX AC ABP48014;



PI Arnaout AM, Li R, Xiong J;  
XX  
DR WPI; 2002-188687/24.  
XX  
XX Novel high affinity integrin polypeptide useful for treating restenosis  
PT and parasitic diseases, comprises all or part of variant integrin alpha  
PT subunit A domain or variant integrin beta subunit A-like domain  
XX  
XX Example 2; Fig 5; 55pp; English.  
XX  
CC The invention relates to a high affinity integrin polypeptide comprising  
CC all or part of a variant integrin alpha subunit A domain or a variant  
CC integrin beta subunit A-like domain. The polypeptide, preferably the  
CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by  
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
CC for determining if a test compound is a candidate compound for binding to  
CC CD1b or for treating an inflammatory disorder, by contacting a test  
CC compound with the polypeptide and determining if the test compound binds  
CC to the polypeptide. The integrin subunits are useful for reducing  
CC skeletal muscle injury, for treating disorders caused by  
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
CC diseases, to purify variant integrin polypeptide ligands and as bait  
CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
CC the human integrin alpha subunit Alpha 11 A domain.  
XX  
SQ Sequence 193 AA;  
XX  
Query Match 1.1%; Score 13; DB 23; Length 193;  
Best Local Similarity 100.0%; Pred. No. 0.00081;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 169 VIVLDGNSNIYPW 181  
DB 8 VIVLDGNSNIYPW 20  
RESULT 24  
AAU76863  
ID AAU76863 standard; Protein; 193 AA.  
XX  
AC AAU76863;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human integrin alpha subunit Alpha 11 variant A domain.  
XX  
KW Human; integrin alpha subunit; A domain; Alpha 11; integrin beta subunit;  
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;  
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;  
KW mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Misc-difference 191  
FT /note= "wild-type Ile substituted by any other amino  
FT acid"  
FT  
XX  
XX WO200209737-A1.  
XX  
XX 07-FEB-2002.  
XX  
XX 31-JUL-2001; 2001WO-US23957.  
XX  
XX 31-JUL-2000; 2000US-221950P.  
XX 11-JAN-2001; 2001US-0758493.  
XX 13-MAR-2001; 2001US-0805354.  
XX  
XX (GEHO ) GEN HOSPITAL CORP.  
XX  
XX

PI Arnaout AM, Li R, Xiong J;  
XX  
DR WPI; 2002-188687/24.  
XX  
XX Novel high affinity integrin polypeptide useful for treating restenosis  
PT and parasitic diseases, comprises all or part of variant integrin alpha  
PT subunit A domain or variant integrin beta subunit A-like domain  
XX  
XX Claim 57; Page -, 55pp; English.  
XX  
CC The invention relates to a high affinity integrin polypeptide comprising  
CC all or part of a variant integrin alpha subunit A domain or a variant  
CC integrin beta subunit A-like domain. The polypeptide, preferably the  
CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by  
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or  
CC V at residue 315 and A at residue 320 have been replaced by C, is useful  
CC for determining if a test compound is a candidate compound for binding to  
CC CD1b or for treating an inflammatory disorder, by contacting a test  
CC compound with the polypeptide and determining if the test compound binds  
CC to the polypeptide. The integrin subunits are useful for reducing  
CC skeletal muscle injury, for treating disorders caused by  
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic  
CC diseases, to purify variant integrin polypeptide ligands and as bait  
CC proteins in two-hybrid or three-hybrid assays. This sequence represents  
CC a human integrin alpha subunit Alpha 11 variant A domain.  
CC Note: This variant sequence is not featured in the specification but has  
CC been derived from the wild-type protein shown in AAU76854.  
XX  
SQ Sequence 193 AA;  
XX  
Query Match 1.1%; Score 13; DB 23; Length 193;  
Best Local Similarity 100.0%; Pred. No. 0.00081;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 169 VIVLDGNSNIYPW 181  
DB 8 VIVLDGNSNIYPW 20  
RESULT 25  
AAU76851  
ID AAU76851 standard; Protein; 195 AA.  
XX  
AC AAU76851;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human integrin alpha subunit Alpha 1 (CD49a) A domain.  
XX  
KW Human; integrin alpha subunit; A domain; CD49a; integrin beta subunit;  
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;  
KW ischaemia-reperfusion injury; immune complex; parasitic disease; Alpha 1;  
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.  
XX  
OS Homo sapiens.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Misc-difference 191  
FT /note= "wild-type Ile substituted by any other amino  
FT acid"  
FT  
XX  
XX WO200209737-A1.  
XX  
XX 07-FEB-2002.  
XX  
XX 31-JUL-2001; 2001WO-US23957.  
XX  
XX 31-JUL-2000; 2000US-221950P.  
XX 11-JAN-2001; 2001US-0758493.  
XX 13-MAR-2001; 2001US-0805354.  
XX  
XX (GEHO ) GEN HOSPITAL CORP.  
XX  
XX Arnaout AM, Li R, Xiong J;  
XX  
XX WPI; 2002-188687/24.  
XX  
XX Novel high affinity integrin polypeptide useful for treating restenosis  
PT





CC blocking antibody capable of binding an epitope of human alpha1 integrin  
CC domain (VLA-1; see AAB50043). The antibody of the present invention is  
CC useful for treating an inflammatory disorder e.g. rheumatoid arthritis,  
CC skin related conditions such as psoriasis, eczema, burns and dermatitis,  
CC asthma, bronchitis, menstrual cramps, tendinitis, bursitis, and the  
CC treatment of pain and headaches, or as an antipyretic, for the treatment  
CC of fever, gastrointestinal conditions such as inflammatory bowel  
CC diseases, Crohn's disease, gastritis and vascular diseases, migraine  
CC headaches, periarthritis nodosa, thyroiditis, aplastic anaemia,  
CC Hodgkin's disease, sarcoidosis, nephrotic syndrome, myocardial  
CC ischaemia, allergic rhinitis, respiratory distress syndrome, endotoxin  
CC shock syndrome and atherosclerosis. The present sequence is the rat  
CC alpha1 integrin domain.  
CC  
SQ Sequence 214 AA;  
Query Match 1.1%; Score 13; DB 22; Length 214;  
Best Local Similarity 100.0%; Pred. No. 0.00089;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 169 VIVLDGNSNIYPM 181  
Db 24 VIVLDGNSNIYPM 36  
RESULT 28  
AAB50042  
ID AAB50042 standard; protein; 214 AA.  
AC AAB50042;  
XX  
XX  
DT 14-MAR-2001 (first entry)  
XX  
DE Human alpha1 integrin domain.  
XX  
KW Human; alpha1 integrin domain; alpha1beta1 function blocking antibody;  
KW inflammatory disorder; rheumatoid arthritis; skin condition; asthma;  
KW bronchitis; headache; antipyretic; fever; gastrointestinal;  
KW vascular disease; autoimmune; respiratory distress syndrome;  
KW endotoxin shock syndrome; atherosclerosis.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200072881-A1.  
XX  
PD 07-DEC-2000.  
XX  
PF 01-JUN-2000; 2000MO-US15004.  
XX  
PR 01-JUN-1999; 99US-0137038.  
PR 29-FEB-2000; 2000US-0185336.  
XX  
XX (BIOJ) BIOGEN INC.  
XX  
PI De Fougereolles A, Gotwals P, Lobb R, Koteliarsky V;  
XX  
XX WPI; 2001-061448/07.  
XX  
PT Use of blocking monoclonal antibody capable of binding to an epitope of  
PT VLA-1 for treating inflammatory disorders, in particular arthritis  
XX  
XX  
PS Example 14; Fig 15; 60pp; English.  
XX  
CC The present invention relates to the use of an alpha1beta1 function  
CC blocking antibody capable of binding an epitope of human alpha1 integrin  
CC domain (VLA-1; see AAB50043). The antibody of the present invention is  
CC useful for treating an inflammatory disorder e.g. rheumatoid arthritis,  
CC skin related conditions such as psoriasis, eczema, burns and dermatitis,  
CC asthma, bronchitis, menstrual cramps, tendinitis, bursitis, and the  
CC treatment of pain and headaches, or as an antipyretic, for the treatment  
CC of fever, gastrointestinal conditions such as inflammatory bowel  
CC diseases, Crohn's disease, gastritis and vascular diseases, migraine

CC headaches, periarthritis nodosa, thyroiditis, aplastic anaemia,  
CC Hodgkin's disease, sarcoidosis, nephrotic syndrome, myocardial  
CC multiple sclerosis, rheumatic fever, type I diabetes, myasthenia  
CC ischaemia, allergic rhinitis, respiratory distress syndrome, endotoxin  
CC shock syndrome and atherosclerosis. The present sequence is the human  
CC alpha1 integrin domain.  
CC  
SQ Sequence 214 AA;  
Query Match 1.1%; Score 13; DB 22; Length 214;  
Best Local Similarity 100.0%; Pred. No. 0.00089;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 169 VIVLDGNSNIYPM 181  
Db 24 VIVLDGNSNIYPM 36  
RESULT 29  
AAU19663  
ID AAU19663 standard; Protein; 707 AA.  
AC AAU19663;  
XX  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Human novel extracellular matrix protein, Seq ID No 313.  
XX  
XX  
KW Human; secreted extracellular matrix protein; immunomodulatory;  
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiac; vascular;  
KW cerebroprotective; thrombolytic; antimicrobial; opthalmic; cytostatic;  
KW antialzheimers; immune/autoimmune disease; HIV infection; anaemia;  
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;  
KW Sezary syndrome; Gaucher's disease; neurological diseases;  
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
KW wound healing; immunogen; gene therapy; antisense; food additive.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200155368-A1.  
XX  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01348.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0234984.  
 PR 27-SEP-2000; 2000US-0235634.  
 PR 27-SEP-2000; 2000US-0235636.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0244675.  
 PR 08-NOV-2000; 2000US-0244676.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251899.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PR XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI, 2001-465572/50.  
 DR N-PSDB; AAS31234.  
 XX  
 PT Nucleic acid molecules encoding human secreted extracellular matrix  
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
 PT Alzheimer's and Parkinson's diseases and cancers -  
 XX  
 PS Claim 11; SEQ ID No 313; 577pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules encoding  
 CC novel human secreted extracellular matrix proteins (SPs). The  
 CC polynucleotides and proteins are used to prevent, treat a medical  
 CC condition in e.g. humans, mice, rabbits, goats, cats, dogs,  
 CC chickens or sheep. For example, disorders associated with decreased  
 CC expression of SPs. The SP polynucleotide or a vector expressing them may  
 CC be administered to treat diseases by gene therapy. Antisense molecules  
 CC may be administered to down regulate expression of SPs by binding with  
 CC the cells own genes and preventing their expression. The polynucleotides  
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
 CC used as antigens to produce antibodies and to identify modulators  
 CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and  
 CC antagonists may also be used to down regulate expression and activity of  
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
 CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease), cardio-/cerebrovascular disorders (e.g. cardiac  
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses



KW Immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;  
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;  
 KW Crohn's disease; nephritis; hyperproliferative disorder;  
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;  
 KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.  
 OS Homo sapiens.  
 PN WO20029435-A1.  
 PD 25-MAY-2000.  
 PF 27-OCT-1999; 99WO-US25031.  
 PR 28-OCT-1998; 98US-0105971.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y,  
 PI Greene JM;  
 DR WPI: 2000-387742/33.  
 XX  
 XX  
 PT Isolated nucleic acid molecules encoding human secreted proteins are  
 PT used for the prevention, amelioration and treatment of autoimmune,  
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,  
 PT wounds, and infectious diseases -  
 PS Claim 1: Page 678-682; 803pp; English.  
 XX  
 XX The present invention relates to 12 secreted human proteins and the  
 CC nucleotide sequences encoding them. The polynucleotide sequences given  
 CC in AAA80606-AB0623 encode the 12 secreted protein sequences given in  
 CC AAB25576-B25593. The human secreted proteins have various activities  
 CC dependent on the tissues in which they are expressed. Examples of the  
 CC activities of the proteins include: immunosuppressant;  
 CC anti-inflammatory; antirheumatic; dermatological;  
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnary;  
 CC antiviral; antibacterial; and antifungal activity. The proteins,  
 CC polypeptides, agonists and antagonists may be used to treat prevent  
 CC and/or diagnose various disease, disorders and conditions examples of  
 CC which include: immune disorders e.g. Addison's disease, rheumatoid  
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders  
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;  
 CC hyperproliferative disorders such as paraproelasma and purpura;  
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;  
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide  
 CC sequences may also be used in wound healing and the treatment of  
 CC infectious diseases. The human secreted protein gene #7 and protein  
 CC sequences are represented in sequences AAA80612 and AAB25582. Secreted  
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences  
 CC AAA80652-AB0661 represent genes which are related to the secreted protein  
 CC gene#7.  
 XX  
 SQ Sequence 1034 AA;  
 QY  
 Db 169 VIPLDGSNSIYPM 181  
 166 VIPLDGSNSIYPM 178  
 Query Match 1.1%; Score 13; DB 21; Length 1034;  
 Best Local Similarity 100.0%; Pred. No. 0.0035;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DE Human Tumour Endothelial Marker polypeptide SEQ ID NO 250.  
 XX  
 XX Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytosclastic;  
 KW normal endothelial marker; pan-endothelial marker; immunostimulant;  
 KW antiangiogenic; tumour; neoangiogenesis; vascularised tumour;  
 KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;  
 KW psoriasis.  
 OS Homo sapiens.  
 PN WO200210217-A2.  
 PD 07-FEB-2002.  
 PF 01-AUG-2001; 2001WO-US24031.  
 PR 02-AUG-2000; 2000US-222599P.  
 PR 11-AUG-2000; 2000US-224360P.  
 PR 11-APR-2001; 2001US-282850P.  
 PA (UYJO ) UNIV JOHNS HOPKINS.  
 PI St Croix B, Kinzler KW, Vogelstein B;  
 DR WPI: 2002-291856/33.  
 DR N-Psdb; ABL92113.  
 XX  
 XX  
 PT An isolated molecule comprising an antibody variable region which  
 PT specifically binds to an extracellular domain of a tumor endothelial  
 PT marker (TEM) protein, useful for inhibiting tumor growth -  
 PS Claim 35; Page 242-245; 331pp; English.  
 XX  
 XX The invention relates to an isolated molecule comprising an antibody  
 CC variable region which specifically binds to an extracellular domain of a  
 CC tumor endothelial marker (TEM) protein selected from ABB90732, ABB90740,  
 CC ABB90749, ABB90789 and ABB90769. The antibodies which bind to TEM  
 CC proteins have cytosclastic, immunostimulant and antiangiogenic activity.  
 CC They are useful for inhibiting tumor growth, neoangiogenesis in  
 CC subjects bearing a vascularised tumour, polycystic kidney disease,  
 CC diabetic retinopathy, rheumatoid arthritis and psoriasis. Human, mouse  
 CC and rat TEM genes and the encoded proteins (ABL92075-ABL92141 and  
 CC ABB90721-ABB90789) are disclosed, as are marker oligonucleotide  
 CC sequences: tumour endothelial markers (TEM) ABL91996-ABL92041 and  
 CC ABL92143-ABL92191; normal endothelial markers (NEM) ABL92042-ABL92074;  
 CC and pan-endothelial markers (PEM) ABL91903-ABL91995.  
 XX  
 SQ Sequence 1179 AA;  
 QY  
 Db 169 VIPLDGSNSIYPM 181  
 174 VIPLDGSNSIYPM 186  
 Query Match 1.1%; Score 13; DB 23; Length 1179;  
 Best Local Similarity 100.0%; Pred. No. 0.0039;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 33  
 ABB90788  
 ID ABB90788 standard; Protein; 1180 AA.  
 XX  
 AC ABB90788;  
 XX  
 DT 30-MAY-2002 (first entry)  
 DE Rat Tumour Endothelial Marker polypeptide SEQ ID NO 307.  
 XX  
 XX Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytosclastic;  
 KW normal endothelial marker; pan-endothelial marker; immunostimulant;  
 KW antiangiogenic; tumour; neoangiogenesis; vascularised tumour;  
 KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;  
 KW psoriasis.

```

XX OS Ratus sp..
XX XX WO200210217-A2.
XX PN
XX PD 07-FEB-2002.
XX XX
XX PF 01-AUG-2001; 2001WO-US24031.
XX XX
XX PR 02-AUG-2000; 2000US-222599P.
XX PR 11-AUG-2000; 2000US-224360P.
XX PR 11-APR-2001; 2001US-282850P.
XX XX
XX PA (UYUO ) UNIV JOHNS HOPKINS.
XX PT
XX PI St Croix B, Kinzler KM, Vogelstein B.
XX XX
XX DR WPI; 2002-291856/33.
XX DR N-PSDB; ABL92141.
XX XX
XX PT An isolated molecule comprising an antibody variable region which
XX specifically binds to an extracellular domain of a tumor endothelial
XX PT marker (TEM) protein, useful for inhibiting tumor growth.
XX PS Disclosure; Page 314-318; 331pp; English.
XX XX
XX CC The invention relates to an isolated molecule comprising an antibody
XX variable region which specifically binds to an extracellular domain of a
XX CC tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,
XX CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM
XX CC proteins have cytostatic, immunostimulant and antiangiogenic activity.
XX CC They are useful for inhibiting tumour growth, neovascularogenesis in
XX CC subjects bearing a vascularised tumour, polycystic kidney disease,
XX CC diabetic retinopathy, rheumatoid arthritis and psoriasis. Human, mouse
XX CC and rat TEM genes and the encoded proteins (ABL92075-ABL92141 and
XX CC ABB90721-ABB90789) are disclosed, as are marker oligonucleotide
XX CC sequences: tumour endothelial markers (TEM) ABL91996-ABL92041 and
XX CC ABL92143-ABL92191; normal endothelial markers (NEM) ABL92042-ABL92074;
XX CC and pan-endothelial markers (PEM) ABL91903-ABL91995.
XX XX
XX SQ Sequence 1180 AA;
XX
XX Query Match 1.1%; Score 13; DB 23; Length 1180;
XX Best Local Similarity 100.0%; Pred. No. 0.0039;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 169 VVLDGNSNIYPM 181
XX |||||
XX DB 174 VVLDGNSNIYPM 186
XX
XX RESULT 34
XX AAU14231
XX ID AAU14231 standard; Protein; 1188 AA.
XX XX
XX AC AAU14231;
XX XX
XX DT 24-OCT-2001 (first entry)
XX XX
XX DE Human novel protein #102.
XX XX
XX KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
XX immunomodulatory; cytosolic; neuroprotective; vulnary; nootropic;
XX anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
XX antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
XX thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
XX Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX tissue regeneration; immune disorder.
XX OS Homo sapiens.
XX XX
XX PN WO200155437-A2.
XX XX

```

```

PD PD 02-AUG-2001.
XX XX
XX PF 25-JAN-2001; 2001WO-US02623.
XX XX
XX PR 25-JAN-2000; 2000US-0491404.
XX XX
XX PA (HYSE-) HYSEQ INC.
XX XX
XX PI Tang YT, Liu C, Dermanac RT;
XX XX
XX DR WPI; 2001-451939/48.
XX DR N-PSDB; AAS22536.
XX XX
XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage -
XX PS Example 4; Page 578-581; 894pp; English.
XX XX
XX CC The invention relates to polynucleotides encoding novel human
XX CC proteins or their active domains. The polypeptides, polynucleotides and
XX CC antibodies raised against the polypeptides are used in a method of
XX CC treatment of a mammal and prevention of disorders caused by the aberrant
XX CC protein expression or activity. The polypeptides can be used as
XX CC molecular weight markers, food supplements, and in antibody production.
XX CC The polypeptides are used to identify compounds which bind to the
XX CC polypeptides. Polynucleotides of the invention are used as probes and
XX CC primers, for sequencing, for chromosome or gene mapping, in the
XX CC production of recombinant proteins, and in generating anti-sense DNA or
XX CC RNA and in gene therapy. Polypeptides of the invention can be used to
XX CC target drugs to a tumour, in assays to determine biological activity, to
XX CC raise antibodies/elicits an immune response, to determine quantitative
XX CC protein levels, as tissue markers, and to isolate receptors or ligands.
XX CC Polypeptides of the invention may also be useful in treating platelet
XX CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX CC ligament and/or nerve tissue, wound healing, treating burns, promoting
XX CC the proliferation, differentiation and survival of stem cells, as a
XX CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX CC fungal infection or from autoimmunity, cancer, allergy, asthma,
XX CC graft-versus-host disease, eczema, haemophilia, thrombosis,
XX CC anti-inflammatory diseases, nervous system disorders, and infection.
XX CC The present sequence represents a protein of the invention.
XX XX
XX SQ Sequence 1188 AA;
XX
XX Query Match 1.1%; Score 13; DB 22; Length 1188;
XX Best Local Similarity 100.0%; Pred. No. 0.004;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 169 VVLDGNSNIYPM 181
XX |||||
XX DB 166 VVLDGNSNIYPM 178
XX
XX RESULT 35.
XX AAU14467
XX ID AAU14467 standard; Protein; 1188 AA.
XX XX
XX AC AAU14467;
XX XX
XX DT 24-OCT-2001 (first entry)
XX XX
XX DE Human novel protein #338.
XX XX
XX KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
XX immunomodulatory; cytosolic; neuroprotective; vulnary; nootropic;
XX anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
XX antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
XX thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
XX Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX tissue regeneration; immune disorder.
XX OS
XX XX
XX PN
XX XX

```

```

OS Homo sapiens.
XX
XX WO200155437-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-US02623.
XX
XX 25-JAN-2000; 2000US-0491404.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-451939/48.
XX
XX N-PSDB; AAS22772.
XX
XX Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage -
XX
XX Example 4; Page 828-831; 894pp; English.
XX
XX The invention relates to polynucleotides encoding novel human
XX proteins or their active domains. The polypeptides, polynucleotides and
XX antibodies raised against the polypeptides are used in a method of
XX treatment of a mammal and prevention of disorders caused by the aberrant
XX protein expression or activity. The polypeptides can be used as
XX molecular weight markers, food supplements, and in antibody production.
XX The polypeptides are used to identify compounds which bind to the
XX polypeptides. Polynucleotides of the invention are used as probes and
XX primers, for sequencing, for chromosome or gene mapping, in the
XX production of recombinant proteins, and in generating anti-sense DNA or
XX RNA and in gene therapy. Polypeptides of the invention can be used to
XX target drugs to a tumour, in assays to determine biological activity, to
XX raise antibodies/ elicit an immune response, to determine quantitative
XX protein levels, as tissue markers, and to isolate receptors or ligands.
XX Polypeptides of the invention may also be useful in treating platelet
XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX ligament and/or nerve tissue, wound healing, treating burns, promoting
XX the proliferation, differentiation and survival of stem cells, as a
XX contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX fungal infection or from autoimmunity, cancer, allergy, asthma,
XX graft-versus-host disease, eczema, haemophilia, thrombosis,
XX anti-inflammatory diseases, nervous system disorders, and infection.
XX The present sequence represents a protein of the invention.
XX
XX Sequence 1188 AA;
XX
XX Query Match 1.1%; Score 13; DB 22; Length 1188;
XX Best Local Similarity 100.0%; Pred. No. 0.004;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 169 VIVLDGNSNIYPW 181
DB 166 VIVLDGNSNIYPW 178

```

RESULT 36  
AAB30929  
ID AAB30929 standard; Protein; 1188 AA.

AC AAB30929;

DT 02-APR-2001 (first entry)

XX Amino acid sequence of a human alpha1 integrin chain.

XX Human; integrin; alpha1 subunit; fibroblast; muscle cell; chondrocyte;  
XX osteoblast; stem cell; cell damage; muscle dystrophy; fibrosis;  
XX wound healing; trauma; rheumatoid arthritis; osteoarthritis;  
XX osteoporosis; cartilage damage; bone damage; cartilage.

```

XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..22
XX Region /note= "signal peptide"
XX Domain 951..972
XX /note= "leucine zipper"
XX 1142..1164
XX /note= "transmembrane domain"
XX
XX WO200075187-A1.
XX
XX 14-DEC-2000.
XX
XX 31-MAY-2000; 2000WO-SE01135.
XX
XX 03-JUN-1999; 99SE-0002056.
XX
XX (ACT1-) ACTIVE BIOTECH AB.
XX
XX Gullberg D;
XX
XX WPI; 2001-071061/08.
XX
XX N-PSDB; AAC86871.
XX
XX Integrin subunit alpha 11 or integrin heterodimer comprising subunit
XX alpha 11 in association with subunit beta, useful for treating muscle
XX dystrophy, fibrosis, trauma, rheumatoid arthritis, and osteoarthritis -
XX
XX Disclosure; Fig 2a-c; 79pp; English.
XX
XX The present sequence represents a human integrin subunit, designated
XX alpha11. The alpha11 polynucleotide and polypeptide are useful as
XX markers of cell target molecules, such as fibroblasts, muscle cells,
XX chondrocytes, osteoblasts, mesenchymally derived cells or stem cells.
XX They are also used for determining the differential-stage of cells
XX during differentiation, development in pathological conditions, in
XX tissue regeneration, in transplantation or in therapeutic and
XX physiological repair of tissues. The pathological conditions involving
XX subunit alpha11 are selected from damage of cells, muscle dystrophy,
XX fibrosis, wound healing, trauma, rheumatoid arthritis, osteoarthritis
XX and osteoporosis, damage of cartilage and bone, and cartilage and bone
XX diseases. The polypeptide is useful for detecting the formation of
XX cartilage during embryonic development, for detecting physiological
XX therapeutic repair of cartilage and muscle, for selection and analysis,
XX or for sorting, isolating or purification of chondrocytes and muscle
XX cells, for detecting regeneration of cartilage or chondrocytes during
XX transplantation of cartilage or chondrocytes during transplantation of
XX cartilage or chondrocytes, respectively, or of muscle or muscle cells
XX during transplantation of muscle or muscle cells, respectively, and for
XX studies of differentiation or chondrocytes or muscle cells.
XX
XX Sequence 1188 AA;
XX
XX Query Match 1.1%; Score 13; DB 22; Length 1188;
XX Best Local Similarity 100.0%; Pred. No. 0.004;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 169 VIVLDGNSNIYPW 181
DB 166 VIVLDGNSNIYPW 178

```

RESULT 37

AAB50085  
ID AAB50085 standard; Protein; 1188 AA.

AC AAB50085;

DT 19-MAR-2001 (first entry)

XX Human A259.

```

XX Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
XX rheumatoid arthritis.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= Signal_peptide
FT Protein 23..1188
FT /label= Mature_protein
FT Domain 1..1141
FT /label= Extracellular_domain
FT Domain 39..74
FT /label= Integrin_alphasubunit_repeat_domain_#1
FT Domain 115..157
FT /label= Integrin_alphasubunit_repeat_domain_#2
FT Domain 164..345
FT /label= I_domain
FT Domain 367..392
FT /label= Integrin_alphasubunit_repeat_domain_#3
FT Domain 421..455
FT /label= Integrin_alphasubunit_repeat_domain_#4
FT Domain 478..516
FT /label= Integrin_alphasubunit_repeat_domain_#5
FT Domain 540..575
FT /label= Integrin_alphasubunit_repeat_domain_#6
FT Domain 602..640
FT /label= Integrin_alphasubunit_repeat_domain_#7
FT Domain 1142..1164
FT /label= Transmembrane_domain
FT Domain 1165..1188
FT /label= Cytoplasmic_domain
XX
PN WO200073339-A1.
XX
PD 07-DEC-2000.
XX
PF 15-MAY-2000; 2000WO-US13262.
XX
PR 28-MAY-1999; 99US-0322790.
PR 27-APR-2000; 2000US-0561263.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Pan Y, Lora JM;
XX
DR WPI; 2001-041142/05.
DR N-PSDB; AAC91901, AAC91902.
XX
PT Nucleic acid encoding alpha-integrin subunits, useful for treatment and
PT diagnosis of fibrosis, e.g. of the liver
XX
PS Claim 8; Fig 1; 164pp; English.
XX
CC The present sequence is human integrin alpha subunit, A259. A259 is
CC homologous with the alpha1 and alpha10 integrin subunits and is
CC overexpressed in fibrosis. A259 is implicated in regulation of
CC proliferation, differentiation and/or function of many different cell
CC types. Inhibitors of A259 activity are useful for the treatment of liver
CC disease, particularly fibrosis, and also fibrosis in other organs
CC (specifically lung and kidney). In addition, A259 can be used for
CC treatment and prevention of cancer, osteoporosis, acute myeloid
CC leukaemia, HIV infection, and rheumatoid arthritis.
XX
SQ Sequence 1188 AA;

```

```

DB 166 VIVLDGNSIYPW 178
|||||
RESULT 38
AAB50087
ID AAB50087 standard; Protein; 1188 AA.
XX
AC AAB50087;
XX
DT 19-MAR-2001 (first entry)
XX
DE Murine A259.
XX
KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
KW rheumatoid arthritis.
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= Signal_peptide
FT Protein 23..1188
FT /label= Mature_protein
FT Domain 1..1141
FT /label= Extracellular_domain
FT Domain 39..74
FT /label= Integrin_alphasubunit_repeat_domain_#1
FT Domain 115..157
FT /label= Integrin_alphasubunit_repeat_domain_#2
FT Domain 164..345
FT /label= I_domain
FT Domain 367..392
FT /label= Integrin_alphasubunit_repeat_domain_#3
FT Domain 421..455
FT /label= Integrin_alphasubunit_repeat_domain_#4
FT Domain 478..516
FT /label= Integrin_alphasubunit_repeat_domain_#5
FT Domain 540..575
FT /label= Integrin_alphasubunit_repeat_domain_#6
FT Domain 602..640
FT /label= Integrin_alphasubunit_repeat_domain_#7
FT Domain 1142..1164
FT /label= Transmembrane_domain
FT Domain 1165..1188
FT /label= Cytoplasmic_domain
XX
PN WO200073339-A1.
XX
PD 07-DEC-2000.
XX
PF 15-MAY-2000; 2000WO-US13262.
XX
PR 28-MAY-1999; 99US-0322790.
PR 27-APR-2000; 2000US-0561263.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Pan Y, Lora JM;
XX
DR WPI; 2001-041142/05.
DR N-PSDB; AAC91904, AAC91905.
XX
PT Nucleic acid encoding alpha-integrin subunits, useful for treatment and
PT diagnosis of fibrosis, e.g. of the liver
XX
PS Claim 8; Fig 5; 164pp; English.
XX
CC The present sequence is murine integrin alpha subunit, A259. A259 is
CC homologous with the alpha1 and alpha10 integrin subunits and is
CC overexpressed in fibrosis. A259 is implicated in regulation of
CC proliferation, differentiation and/or function of many different cell

```



types. Inhibitors of A259 activity are useful for the treatment of liver disease, particularly fibrosis, and also fibrosis in other organs (specifically lung and kidney). In addition, A259 can be used for treatment and prevention of cancer, osteoporosis, acute myeloid leukaemia, HIV infection, and rheumatoid arthritis.

Sequence 1188 AA;

Query Match 1.1%; Score 13; DB 22; Length 1188;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 VIVLDGNSNIYPW 181  
166 VIVLDGNSNIYPW 178

RESULT 39

ID AAU10551 standard; Protein; 1188 AA.

AC AAU10551;

DI 14-FEB-2002 (first entry)

DE Human A259 polypeptide.

Human; A259: integrin alpha subunit; integrin alpha 10; secreted protein; liver disease; fibrosis; lung; kidney; bone associated disorder; blood; cartilage associated disorder; haematopoietic disorder; bone marrow; immune related disease; apoptotic disorder; neuronal tissue disease; neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic; nephrotropic; immunomodulator; anti-inflammatory; neuroprotective; antiarrhythmic; antinaemic; antiallergic; antiasthmatic; dermatological; antidiabetic; anticonvulsant; antiparkinsonian.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..22 /note= "Signal peptide"

FT Domain 1..1141 /note= "Extracellular domain"

FT Protein 23..1188 /note= "Mature human A259"

FT Domain 37..90 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"

FT Domain 115..157 /note= "Integrin alpha repeat domain"

FT Domain 164..345 /note= "I domain or Von Willebrand Factor type A domain"

FT Domain 367..392 /note= "Integrin alpha repeat domain"

FT Domain 421..472 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"

FT Domain 476..532 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"

FT Domain 538..593 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"

FT Domain 600..654 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"

FT Domain 1142..1164 /note= "Transmembrane domain"

FT Domain 1165..1188 /note= "Cytoplasmic domain"

FT Domain 1165..1188 /note= "Cytoplasmic domain"

FT Domain 1165..1188 /note= "Cytoplasmic domain"

FT Domain 1165..1188 /note= "Cytoplasmic domain"

XX WO200181414-A2.

PD 01-NOV-2001.

XX 27-APR-2001; 2001WO-US13516.

XX 27-APR-2000; 2000US-0561263.

XX (MILL-) MILLENNIUM PHARM INC.

XX Pan Y, Lora J;

XX WPI; 2002-041397/05.

XX N-PSDB; AAS16873.

PT New A259 nucleic acids and polypeptides, which comprise integrin alpha subunit, useful for diagnosing, preventing or treating e.g. liver disease, kidney or lung fibrosis, cancers, blood disorders or immune related diseases

PS Claim 9; Fig 1; 168pp; English.

The invention relates to human and murine A259 nucleic acid molecules which encode secreted proteins with homology to integrin alpha subunits, specifically to integrin alpha 10. The A259 polypeptide and nucleic acid are useful for treating liver disease or fibrosis, particularly kidney fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also useful for diagnosing, preventing or treating cartilage and bone associated disorders (such as bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and osteoporosis), bone marrow, blood and haematopoietic disorders (such as acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune related diseases (such as HIV, viral infections, cancers, T cell autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g. asthma and psoriasis), apoptotic disorders (such as systemic lupus erythematosus and insulin-dependent diabetes mellitus), diseases of the neuronal tissues (such as epilepsy and muscular dystrophy) and neurodegenerative diseases (such as Parkinson's disease and Huntington's disease). This sequence represents the human A259 polypeptide.

Sequence 1188 AA;

Query Match 1.1%; Score 13; DB 23; Length 1188;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 VIVLDGNSNIYPW 181

Db 166 VIVLDGNSNIYPW 178

RESULT 40

ID AAU10552 standard; Protein; 1188 AA.

AC AAU10552;

DI 14-FEB-2002 (first entry)

DE Murine A259 polypeptide.

Mouse; A259: integrin alpha subunit; integrin alpha 10; secreted protein; liver disease; fibrosis; lung; kidney; bone associated disorder; blood; cartilage associated disorder; haematopoietic disorder; bone marrow; immune related disease; apoptotic disorder; neuronal tissue disease; neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic; nephrotropic; immunomodulator; anti-inflammatory; neuroprotective; antiarrhythmic; antinaemic; antiallergic; antiasthmatic; dermatological; antidiabetic; anticonvulsant; antiparkinsonian.

OS Mus musculus.

XX Key Location/Qualifiers

FT Peptide 1..22 /note= "Signal peptide"

Accession	Protein	Domain	Score	DB	Length	Mismatches	Indels	Gaps
FT	Domain	1..1141						
FT	/note="Extracellular domain"							
FT	Protein	23..1188						
FT	/note="Mature murine A259"							
FT	Domain	39..74						
FT	/note="Integrin alpha repeat domain"							
FT	Domain	115..157						
FT	/note="Integrin alpha repeat domain"							
FT	Domain	164..345						
FT	/note="Integrin alpha repeat domain"							
FT	Domain	367..392						
FT	/note="Integrin alpha repeat domain"							
FT	Domain	421..455						
FT	/note="Integrin alpha repeat domain"							
FT	Domain	478..516						
FT	/note="Integrin alpha repeat domain"							
FT	Domain	540..575						
FT	/note="Integrin alpha repeat domain"							
FT	Domain	602..640						
FT	/note="Integrin alpha repeat domain"							
FT	Domain	1142..1164						
FT	/note="Transmembrane domain"							
FT	Domain	1165..1188						
FT	/note="Cytoplasmic domain"							
PD	MO200181414-A2.							
PD	01-NOV-2001.							
XX	27-APR-2001; 2001WO-US13516.							
XX	27-APR-2000; 2000US-0561263.							
XX	(MILL-) MILLENNIUM PHARM INC.							
XX	Pan Y, Lora J;							
XX	WPI; 2002-041397/05.							
XX	DR N-PSDB; AAS16874.							
XX	New A259 nucleic acids and polypeptides, which comprise integrin alpha subunit, useful for diagnosing, preventing or treating e.g. liver disease, kidney or lung fibrosis, cancers, blood disorders or immune related diseases							
XX	Claim 9; Fig 5; 168p; English.							
XX	The invention relates to human and murine A259 nucleic acid molecules which encode secreted proteins with homology to integrin alpha subunits, specifically to integrin alpha 10. The A259 polypeptide and nucleic acid are useful for treating liver disease or fibrosis, particularly kidney fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also useful for diagnosing, preventing or treating cartilage and bone associated disorders (such as bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and osteoporosis), bone marrow, blood and haematopoietic disorders (such as acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune related diseases (such as HIV, viral infections, cancers, T cell autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g. asthma and psoriasis), apoptotic disorders (such as systemic lupus erythematosus and insulin-dependent diabetes mellitus), diseases of the neuronal tissues (such as epilepsy and muscular dystrophy) and neurodegenerative diseases (such as Parkinson's disease and Huntington's disease). This sequence represents the murine A259 polypeptide.							
XX	Sequence 1188 AA;							
XX	Query Match 1.1%; Score 13; DB 23; Length 1188;							
XX	Best Local Similarity 100.0%; Pred. No. 0.004;							
XX	Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0							

DB	166	VITLDSGNS1PFW	178
RESULT 41			
AA025582			
ID	AA025582	standard; Protein; 1189 AA.	
XX			
AC	AA025582;		
XX			
DT	21-NOV-2000	(first entry)	
XX			
DE	ITGA11 protein encoded by human secreted protein gene #7.		
XX			
KM	Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;		
KM	antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;		
KM	antitumor; vulnery; antiviral; antibacterial; antifungal;		
KM	immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;		
KM	multiple sclerosis; inflammatory disorder; inflammatory bowel disease;		
KM	Crohn's disease; nephritis; hyperproliferative disorder;		
KM	cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;		
KM	melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.		
XX			
OS	Homo sapiens.		
XX			
PN	W0200029435-A1.		
XX			
PD	25-MAY-2000.		
XX			
PF	27-OCT-1999; 99MO-US25031.		
XX			
PR	28-OCT-1998; 98US-0105971.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y,		
XX	Greene JW,		
DR	WPI: 2000-387742/33.		
XX	N-PSDB: AAA060612.		
XX			
PT	Isolated nucleic acid molecules encoding human secreted proteins are		
XX	used for the prevention, amelioration and treatment of autoimmune,		
XX	inflammatory, hyperproliferative and cardiovascular disorders, cancer,		
XX	wounds, and infectious diseases -		
XX			
PS	Claim 1; Figure 19A-F; 803pp; English.		
XX			
CC	The present invention relates to 12 secreted human proteins and the		
CC	nucleotide sequences encoding them. The polynucleotide sequences given		
CC	in AA060612-80623 encode the 12 secreted protein sequences given in		
CC	AA025576-825593. The human secreted proteins have various activities		
CC	dependent on the tissues in which they are expressed. Examples of the		
CC	activities of the proteins include: immunosuppressant;		
CC	anti-inflammatory; antiarthritic; antirheumatic, dermatological;		
CC	antiproliferative; antiarteriosclerotic; anticancer; vulnery;		
CC	antiviral; antibacterial; and antifungal activity. The proteins,		
CC	polypeptides, agonists and antagonists may be used to treat prevent		
CC	and/or diagnose various disease, disorders and conditions examples of		
CC	which include: immune disorders e.g. Addison's disease, rheumatoid		
CC	arthritis, dermatitis, and multiple sclerosis; inflammatory disorders		
CC	e.g. inflammatory bowel disease, Crohn's disease and nephritis;		
CC	hyperproliferative disorders such as paraproteinemias and purpura;		
CC	cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;		
CC	cancer e.g. melanoma and lymphoma. The proteins and polynucleotide		
CC	sequences may also be used in wound healing and the treatment of		
CC	infectious diseases. The human secreted protein gene #7 and protein		
CC	sequences are represented in sequences AA060612 and AA025582. Secreted		
CC	protein gene #7 is located at position chromosome 15 q22.3-23. Sequences		
CC	AA060612-80661 represent genes which are related to the secreted protein		
CC	gene#7.		
XX			
XX			
XX			
Sequence	1189 AA;		

Query Match 1.1%; Score 13; DB 21; Length 1189;  
 Best Local Similarity 100.0%; Pred. No. 0.004;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181  
 |||||  
 Db 166 VIVLDGNSIYPM 178

## RESULT 42

ABG12949  
 ID ABG12949 standard; Protein; 1189 AA.

AC ABG12949;  
 XX

DT 13-FEB-2002 (first entry)  
 XX

DE Novel human diagnostic protein #12940.  
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX

OS Homo sapiens.  
 XX

PN WO200175067-A2.  
 XX

PD 11-OCT-2001.  
 XX

PF 30-MAR-2001; 2001WO-US08631.  
 XX

PR 31-MAR-2000; 2000US-0540217.  
 XX

PR 23-AUG-2000; 2000US-0649167.  
 XX

PA (HYSE-) HYSEQ INC.  
 XX

PI Drmanac RT, Liu C, Tang YT;  
 XX

DR WPI; 2001-639362/73.  
 XX

DR N-PSDB; AAS77136.  
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PS Claim 20; SEQ ID No 4308; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG0377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 1189 AA;

Query Match 1.1%; Score 13; DB 22; Length 1189;  
 Best Local Similarity 100.0%; Pred. No. 0.004;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181  
 |||||  
 Db 166 VIVLDGNSIYPM 178

## RESULT 43

AAV07728  
 ID AAV07728 standard; protein; 1183 AA.

AC AAV07728;  
 XX

DT 01-JUL-1999 (first entry)  
 XX

DE Armenian hamster alpha-1 integrin subunit protein.  
 XX

KW VEGF; tumour angiogenesis inhibition; vascular endothelial growth factor;  
 KW integrin cell surface receptor; capillary; blood vessel; hamster;  
 KW alpha-1 subunit; alpha-2 subunit.  
 XX

OS Cricetus migratorius.  
 XX

PN WO916465-A1.  
 XX

PD 08-APR-1999.  
 XX

PF 30-SEP-1997; 97WO-US17485.  
 XX

PR 30-SEP-1997; 97WO-US17485.  
 XX

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX

PA (CLAF/) CLAFREY K P.  
 XX

PA (DETM/) DETMAR M.  
 XX

PA (SENG/) SENGER D R.  
 XX

PI Clafey KP, Detmar M, Senger DR;  
 XX

DR WPI; 1999-254930/21.  
 XX

PT Inhibition of tumor angiogenesis through interaction of vascular  
 PT endothelial growth factor and integrin cell surface receptors

PS Disclosure; Fig 2A-C; 64pp; English.

XX This invention describes a novel method for the inhibition of tumour  
 CC angiogenesis mediated by vascular endothelial growth factor (VEGF) and  
 CC integrin cell surface receptors expressed in vasculature of living  
 CC subjects. The method inhibits new capillary and new blood vessel  
 CC formation both within a tumour mass itself as well as in the immediately  
 CC adjacent blood vasculature surrounding the perimeter of the tumour mass.  
 CC Interaction and dependence upon VEGF to induce specific integrin  
 CC heterodimers in tumour angiogenesis provides a novel method for  
 CC inhibiting tumour angiogenesis, and unlike prior art relies on the  
 CC specific inter-relationship of VEGF and integrins, rather than  
 CC concentrating solely on one specific class of protein.

SO Sequence 1183 AA;

Query Match 1.0%; Score 12; DB 20; Length 1183;  
 Best Local Similarity 100.0%; Pred. No. 0.037;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 180  
 |||||  
 Db 176 VIVLDGNSIYPM 187

## RESULT 44

ABG12950  
 ID ABG12950 standard; Protein; 117 AA.

AC ABG12950;  
 XX

XX 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #12941.  
 XX  
 KM Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT.  
 XX  
 DR WPI; 2001-639362/73.  
 XX N-PSDB; AAS77137.  
 DR  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnosis, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 43309; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes.  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. Abg00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 117 AA;  
 XX  
 Query Match 0.9%; Score 11; DB 22; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 0.046;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLALLLV 1141  
 |||||  
 DB 62 LGGLLALLLV 72

RESULT 45  
 AAU09125  
 ID AAU09125 standard; protein; 185 AA.  
 XX  
 AC AAU09125;  
 XX  
 DT 16-JAN-2002 (first entry)

XX Human integrin alpha2 I domain.  
 DE  
 XX  
 KM Human: integrin alpha2; I domain; protein co-ordinate data; thrombolytic;  
 KM cardiant; cerebroprotective; hepatotropic; anti-inflammatory; vitucide;  
 KM vasotrophic; antiarteriosclerotic; thrombotic disorder.  
 KM myocardial infarction; stroke; acute thrombosis; angioplasty;  
 KM coronary bypass grafting; liver fibrosis; liver necrosis; hepatitis;  
 KM arterial occlusion; restenosis; atherogenesis; anti-platelet therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200173444-A2.  
 PD 04-OCT-2001.  
 XX  
 PF 27-MAR-2001; 2001WO-GB01358.  
 XX  
 PR 28-MAR-2000; 2000US-192180P.  
 XX  
 PA (UYCA-) UNIV CAMBRIDGE TECH SERVICES LTD.  
 XX  
 PI Farndale RW, Emley J, Knight CG, Barnes MJ, Liddington RC;  
 XX  
 DR WPI; 2001-648466/74.  
 XX  
 PT New methods for identifying inhibitors of an I-domain-containing  
 PT polypeptide, particularly integrin, comprises use of coordinates of  
 PT peptide/receptor crystal structure -  
 XX  
 PS Example: Page 47; 100bp; English.  
 XX  
 CC The invention relates to methods for identifying a potential inhibitor of  
 CC an I-domain comprising polypeptide, particularly integrin, comprising  
 CC using a three dimensional structure of the integrin alpha2 I-domain as  
 CC defined in the specification to design or select a potential inhibitor.  
 CC Identified inhibitors are used to treat a disorder or disease. Such  
 CC disorders include thrombotic disorders, myocardial infarction and stroke,  
 CC acute thrombosis associated with angioplasty and coronary bypass  
 CC grafting, and with liver fibrosis or thrombotic complication of liver  
 CC necrosis after hepatitis. Inhibition of platelet alpha2beta1 may be used  
 CC to treat longer term occlusion of arteries, restenosis and atherogenesis.  
 CC Collagen receptor antagonism may be used as anti-platelet therapy.  
 CC The present sequence is the integrin alpha2 I domain.  
 XX  
 SQ Sequence 185 AA;  
 XX  
 Query Match 0.9%; Score 11; DB 22; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHSDG 282  
 |||||  
 DB 109 VVVTGESHSDG 119

RESULT 46  
 ABB72288  
 ID ABB72288 standard; Protein; 545 AA.  
 XX  
 AC ABB72288;  
 XX  
 DT 04-APR-2002 (first entry)

XX Murine protein isolated from skin cells SEQ ID NO: 500.  
 XX  
 DE Human; rat; mouse; skin cell; skin wound; cancer; growth defect;  
 XX

KW		developmental defect; inflammatory disease; dermatological; vulnery;
KV		immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX	Mus sp.	
OS	WO200190357-A1.	
PN		
XX	29-NOV-2001.	
PD		
XX		
PF	24-MAY-2001; 2001WO-NZ00099.	
XX		
PR	24-MAY-2000; 2000US-2066SOP.	
PR	25-JUL-2000; 2000US-221232P.	
XX	(GENE-) GENESIS RES & DEV CORP LTD.	
PA		
PI	Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;	
XX		
DR	WPI; 2002-122020/16.	
XX		
PT	New polynucleotides and polypeptides encoded by the polynucleotides	
PT	isolated from skin cells, useful for treating skin wounds, cancers,	
PT	growth and developmental defects, inflammatory diseases, or for	
PT	modulating immune responses -	
PS		
XX	Claim 4; Page 305-306; 466pp; English.	
CC	The present invention provides the protein and coding sequences of cDNAs	
CC	isolated from human, murine and rat skin cell libraries. The sequences	
CC	can be used in the development of therapeutic agents useful in the	
CC	treatment of skin diseases, including skin wounds, cancer, growth	
CC	defects, developmental defects and inflammatory diseases. The proteins	
CC	have important roles in the induction of hair growth, cell proliferation	
CC	and cell-cell interaction, in maintaining tissue integrity, in wound	
CC	healing and in modulating immune responses. The present sequence is a	
XX	polypeptide of the invention.	
SQ	Sequence 545 AA:	
	Query Match 0.9%; Score 11; DB 23; Length 545;	
	Best Local Similarity 100.0%; Pred. No. 0.18;	
	Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	1131 LGULLLALLV 1141           	
DB	507 LGULLLALLV 517	
RESULT 47		
ABB72300		
ID	ABB72300 standard; Protein; 688 AA.	
XX		
AC	ABB72300;	
XX		
DT	04-APR-2002 (first entry)	
XX		
DE	Rat protein isolated from skin cells SEQ ID NO: 624.	
XX		
KW	Human; rat; mouse; skin cell; skin wound; cancer; growth defect;	
KW	developmental defect; inflammatory disease; dermatological; vulnery;	
KW	immunomodulator; anti-inflammatory; cytostatic; neuroprotective.	
XX		
OS	Rattus sp.	
XX		
PN	WO200190357-A1.	
XX		
PD	29-NOV-2001.	
XX		
PF	24-MAY-2001; 2001WO-NZ00099.	
XX		
PR	24-MAY-2000; 2000US-2066SOP.	
PR	25-JUL-2000; 2000US-221232P.	
XX		

PA	(GENE-) GENESIS RES & DEV CORP LTD.
XX	
PI	Watson JD, Strachan L, Sleeman M, Orrust R, Murison JG, Kumble KD,
XX	
DR	WPI; 2002-122020/16.
DR	N-PSDB; ABL34985.
XX	
PT	New polynucleotides and polypeptides encoded by the polynucleotides
PT	isolated from skin cells, useful for treating skin wounds, cancers,
PT	growth and developmental defects, inflammatory diseases, or for
PT	modulating immune responses
XX	
PS	Claim 4; Page 388-390; 466BP; English.
XX	
CC	The present invention provides the protein and coding sequences of cDNAs
CC	isolated from human, murine and rat skin cell libraries. The sequences
CC	can be used in the development of therapeutic agents useful in the
CC	treatment of skin diseases, including skin wounds, cancer, growth
CC	defects, developmental defects and inflammatory diseases. The proteins
CC	have important roles in the induction of hair growth, cell proliferation
CC	and cell-cell interaction, in maintaining tissue integrity, in wound
CC	healing and in modulating immune responses. The present sequence is a
CC	polypeptide of the invention.
XX	
SO	Sequence 688 AA;
XX	
Query Match	0.9%; Score 11; DB 23; Length 688;
Best Local Similarity	100.0%; Pred. No. 0.22;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1131 LGGILLALLV 1141
DB	650 LGGILLALLV 660
XX	
RESULT 48	
ABB72289	
ID	ABB72289 standard; Protein; 696 AA.
XX	
AC	ABB72289;
XX	
DT	04-APR-2002 (first entry)
DE	Rat protein isolated from skin cells SEQ ID NO: 501.
XX	
KW	Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW	developmental defect; inflammatory disease; dermatological; vulnary;
KW	immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX	
OS	Rattus sp.
XX	
PN	WO200190357-A1.
PD	29-NOV-2001.
XX	
PF	24-MAY-2001; 2001WO-NZ00039.
XX	
PR	24-MAY-2000; 2000US-206650P.
PR	25-JUL-2000; 2000US-221232P.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
PI	Watson JD, Strachan L, Sleeman M, Orrust R, Murison JG, Kumble KD,
XX	
DR	WPI; 2002-122020/16.
XX	
PT	New polynucleotides and polypeptides encoded by the polynucleotides
PT	isolated from skin cells, useful for treating skin wounds, cancers,
PT	growth and developmental defects, inflammatory diseases, or for
PT	modulating immune responses
XX	
PS	Claim 4; Page 306-307; 466BP; English.

CC The present invention provides the protein and coding sequences of cDNAs  
 CC isolated from human, murine and rat skin cell libraries. The sequences  
 CC can be used in the development of therapeutic agents useful in the  
 CC treatment of skin diseases, including skin wounds, cancer, growth  
 CC defects, developmental defects and inflammatory diseases. The proteins  
 CC have important roles in the induction of hair growth, cell proliferation  
 CC and cell-cell interaction, in maintaining tissue integrity, in wound  
 CC healing and in modulating immune responses. The present sequence is a  
 CC polypeptide of the invention.

XX  
 SQ Sequence 696 AA:

Query Match 0.9%; Score 11; DB 23; Length 696;  
 Best Local Similarity 100.0%; Pred. No. 0.22;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGLLLALLLV 1141  
 DB 650 LGLLLALLLV 660

RESULT 49

AAW98238  
 ID AAW98238 standard; Protein; 697 AA.

AC AAW98238;

DT 31-MAR-1999 (first entry)

DE H. pylori GHP0 702 protein.

KM GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KM peptic ulcer disease.

OS Helicobacter pylori.

PN WO9843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-US06371.

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

PA (HUMA-) HUMAN GENOME SCI INC.

PI (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PI A1-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;

DR WPI; 1998-542293/46.

DR N-PSDB; AAX13957.

PT New isolated Helicobacter polynucleotides - used to develop products

PT for the diagnosis, prevention and treatment of Helicobacter

PT infections and gastrointestinal diseases

PS Claim 8; Page 192-195; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHP0 protein of the

XX invention. The polypeptides can be used for preventing or treating

XX Helicobacter infections, and gastroduodenal diseases associated with

XX CC these infections, including acute, chronic, and atrophic gastritis, and

XX CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be

XX CC used for the production of antibodies. The products can also be used for

XX SQ Sequence 697 AA;

Query Match 0.9%; Score 11; DB 19; Length 697;  
 Best Local Similarity 100.0%; Pred. No. 0.22;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 LMLIGSVLGG 1133  
 DB 128 LMLIGSVLGG 138

RESULT 50

ABG25584  
 ID ABG25584 standard; Protein; 823 AA.

AC ABG25584;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #25575.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM Food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS89771.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

PS Claim 20; SEQ ID No 55943; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX CC polynucleotide (II) sequences. (I) is useful as hybridisation probes,

XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX CC and gene mapping, and in recombinant production of (II). The

XX CC polynucleotides are also used in diagnostics as expressed sequence tags

XX CC for identifying expressed genes. (I) is useful in gene therapy techniques

XX CC to restore normal activity of (II) or to treat disease states involving

XX CC (II). (II) is useful for generating antibodies against it, detecting or

XX CC quantitating a polypeptide in tissue, as molecular weight markers and as

XX CC a food supplement. (II) and its binding partners are useful in medical

XX CC imaging of sites expressing (II). (I) and (II) are useful for treating

XX CC disorders involving aberrant protein expression or biological activity.

XX CC The polypeptide and polynucleotide sequences have applications in

XX CC diagnostics, forensics, gene mapping, identification of mutations

XX SQ Sequence 823 AA;

Query Match 0.9%; Score 11; DB 22; Length 823;  
 Best Local Similarity 100.0%; Pred. No. 0.25;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHG 282

Db 191 VVVTGESHHDG 201

## RESULT 51

ABG29239  
ID ABG29239 standard; Protein; 979 AA.  
XX  
XX ABG29239;  
XX  
XX 18-FEB-2002 (first entry)  
XX  
XX  
XX Novel human diagnostic protein #29230.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS93426.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 20; SEQ ID No 59598; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Sequence 979 AA;

Query Match 0.9%; Score 11; DB 22; Length 979;

Best Local Similarity 100.0%; Pred. No. 0.3;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282

Db 77 VVVTGESHHDG 87

## RESULT 52

AAW07729  
ID AAW07729 standard; protein; 1183 AA.  
XX  
XX AAW07729;  
XX  
XX 01-JUL-1999 (first entry)  
XX  
XX  
XX Armenian hamster alpha-2 integrin subunit protein.  
XX  
XX VEGF; tumour angiogenesis inhibitor; vascular endothelial growth factor;  
KM integrin cell surface receptor; capillary; blood vessel; hamster;  
KM alpha-1 subunit; alpha-2 subunit.  
XX  
XX Cricetus migratorius.  
XX  
XX WO9916465-A1.  
XX  
XX 08-APR-1999.  
XX  
XX 30-SEP-1997; 97WO-US17485.  
XX  
XX 30-SEP-1997; 97WO-US17485.  
XX  
XX 30-SEP-1997; 97WO-US17485.  
XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX (CLAF/) CLAFFEY K P.  
XX (DETM/) DETMAR M.  
XX (SENG/) SENG D R.  
XX  
XX Claffey KP, Detmar M, Senger DR;  
XX  
XX WPI; 1999-254930/21.  
XX  
XX Inhibition of tumor angiogenesis through interaction of vascular  
PT endothelial growth factor and integrin cell surface receptors  
PT  
XX  
XX Disclosure; Fig 3A-C; 64pp; English.  
XX  
XX This invention describes a novel method for the inhibition of tumour  
CC angiogenesis mediated by vascular endothelial growth factor (VEGF) and  
CC integrin cell surface receptors expressed in vasculature of living  
CC subjects. The method inhibits new capillary and new blood vessel  
CC formation both within a tumour mass itself as well as in the immediately  
CC adjacent blood vasculature surrounding the perimeter of the tumour mass.  
CC Interaction and dependence upon VEGF to induce specific integrin  
CC heterodimers in tumour angiogenesis provides a novel method for  
CC inhibiting tumour angiogenesis, and unlike prior art relies on the  
CC specific inter-relationship of VEGF and integrins, rather than  
CC concentrating solely on one specific class of protein.  
XX

Sequence 1183 AA;

Query Match 0.9%; Score 11; DB 20; Length 1183;

Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282

Db 279 VVVTGESHHDG 289

## RESULT 53

AAW70542  
ID AAW70542 standard; Protein; 1367 AA.

AAW70542;

26-JAN-1999 (first entry)

Integrin alpha-2 chain.

XX

KW Integrin; alpha-4 chain; immunoglobulin; chimeric; heterodimer complex;  
 KW inhibitor; binding; ligand; blood platelet; hemostatic; diagnostic agent;  
 XX human.  
 OS Homo sapiens.  
 XX  
 XX  
 FT Key Location/Qualifiers  
 FT Peptide 1..29  
 FT /note="signal peptide"  
 FT Protein 30..1367  
 FT /note="mature protein"  
 PN WO9832771-A1.  
 XX  
 XX 30-JUL-1998.  
 PD  
 XX  
 XX 29-JAN-1998; 98WO-JD00370.  
 PF  
 XX 29-AUG-1997; 97JP-0234544.  
 PR 29-JAN-1997; 97JP-0015118.  
 XX  
 XX (TORA ) TORAY IND INC.  
 PA  
 XX Kainoh M, Tanaka T;  
 PI  
 DR WPI: 1998-427881/36.  
 DR N-PSDB; AAV33774.  
 XX  
 XX Integrin-immunoglobulin chimeric protein heterodimer complexes as  
 PT platelet substitutes - contain the alpha and beta integrin chains  
 PT associated in stable state and bind to extracellular matrix in the  
 PT presence of plasma components  
 XX  
 PS Claim 8; Pages 62-71; 87pp; Japanese.  
 XX  
 CC This represents an integrin alpha-2 chain. The invention provides  
 CC integrin-immunoglobulin chimeric protein heterodimer complexes that  
 CC comprise an integrin alpha or beta chain associated with an  
 CC immunoglobulin light or heavy chain. These chimeric proteins form  
 CC heterodimer complexes, in particular with a chimeric protein containing  
 CC an integrin alpha chain and an immunoglobulin chain with a chimeric  
 CC protein containing an integrin beta chain and an immunoglobulin chain;  
 CC the immunoglobulin chain in each case may be a heavy chain, or one of the  
 CC two may be a light chain. The integrin alpha chain is preferably alpha 4  
 CC or alpha 2 and the integrin beta chain is preferably beta 1. Animal cells  
 CC transformed with vectors containing the DNA coding for the above chimeric  
 CC proteins can be used in the preparation of the chimeric proteins and  
 CC their heterodimer complexes. The heterodimer complexes, which are useful  
 CC for testing potential promoters and inhibitors of the binding of  
 CC integrins to their ligands, function as blood platelet substitutes and  
 CC hemostatics and as diagnostic agents.  
 CC  
 SQ Sequence 1367 AA;  
 QY  
 Query Match 0.9%; Score 11; DB 19; Length 1367;  
 Best Local Similarity 100.0%; Pred. No. 0.4;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

KW Human mucosal lymphocyte-1 antigen; HML-1; integrin;  
 KW cell attachment; T-cell.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 FT Key Location/Qualifiers  
 FT Peptide 1..18  
 FT /label= Sig\_peptide  
 FT Domain 144..198  
 FT /label= X-domain  
 FT Cleavage-site 177..178  
 FT /label= Proteolytic\_cleavage\_site  
 FT Region 180..197  
 FT /label= Highly\_charged\_region  
 FT Region 1124..1146  
 FT /label= Transmembrane\_region  
 FT Region 1147..1178  
 FT /label= Cytoplasmic\_tail  
 PN WO9522610-A1.  
 XX  
 XX 24-AUG-1995.  
 PD  
 XX  
 XX 15-FEB-1995; 95WO-US02044.  
 PF  
 XX 18-FEB-1994; 94US-0199776.  
 PR  
 XX (BIGHM ) BRIGHAM & WOMENS HOSPITAL.  
 PA  
 XX Brenner MB, Parker CM;  
 PI  
 DR WPI: 1995-302716/39.  
 DR N-PSDB; AAT03885.  
 XX  
 XX New isolated integrin alpha sub-unit and peptide(s) - exhibit cell  
 PT attachment activity or block activity of intra-epithelial  
 PT lymphocytes  
 PT  
 PS Claim 4; Page 47-51; 75pp; English.  
 XX  
 CC The human mucosal lymphocyte-1 alpha-E chain (AAR82656) encoded by a  
 CC cDNA clone (AAT03885) isolated from an intra-epithelial lymphocyte  
 CC cDNA library contains an 'X domain' (AAR82635) unique to integrins,  
 CC believed to be involved in the adhesion of T-cells to epithelial  
 CC cells in vivo.  
 CC  
 SQ Sequence 1178 AA;  
 QY  
 Query Match 0.9%; Score 10; DB 16; Length 1178;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 54  
 AAR82656  
 ID AAR82656 standard; Protein; 1178 AA.  
 XX  
 AC AAR82656;  
 XX  
 DT 25-FEB-1996 (first entry)  
 XX  
 DE Human mucosal lymphocyte-1 alpha-E chain.  
 XX

RESULT 55  
 ABG05837  
 ID ABG05837 standard; Protein; 1223 AA.  
 XX  
 AC ABG05837;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #5828.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2;



```

XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS70024.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostic, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
XX Claim 20; SEQ ID No 36196; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1223 AA;

```

```

Query Match 0.9%; Score 10; DB 22; Length 1223;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 483 GSYFSELCF 492
   |||||
Db 552 GSYFSELCF 561

```

```

RESULT 56
AAV32249
ID AAV32249 standard; peptide; 13 AA.
XX
XX AAV32249;
XX
XX 15-FEB-2000 (first entry)
XX
XX Bovine integrin subunit alpha-10 cryptic peptide 5.
XX
XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
XX osteoarthritis; osteoarthritis; cancer; atherosclerosis;
XX inflammation; therapy; cartilage; chondrocyte; osteoblast;
XX fibroblast; vaccine; marker.
XX
XX Bos taurus.
XX
XX Key Location/Qualifiers
FH

```

```

FT Misc-difference 10
FT /note= "unidentified"
XX
XX WO9951639-A1.
XX
XX 14-OCT-1999.
XX
XX 31-MAR-1999; 99WO-SE00544.
XX
XX 02-APR-1998; 98SE-0001164.
XX
XX 28-JAN-1999; 99SE-0000319.
XX
XX (ACTI-) ACTIVE BIOTECH AB.
XX
XX Lundgren-Akerlund E;
XX
XX WPI; 2000-052639/04.
XX
XX New isolated integrin subunit alpha-10, used as a marker or target
XX PT molecule for cells during development, regeneration and pathological
XX PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
XX PT inflammation -
XX
XX Example 2; Fig 2; 90pp; English.
XX
XX This sequence represents tryptic peptide 5 of integrin subunit
XX CC alpha-10 (ISa10) isolated from bovine chondrocytes. 6 tryptic
XX CC peptides (see AAV32245-50) were obtained. PCR primers (see
XX CC AAV324721-24) based on peptides 1 and 2 were used to obtain bovine
XX CC ISa10 cDNA fragments. These were used to screen a human
XX CC articular chondrocyte cDNA library in order to isolate human ISa10
XX CC cDNA (see AAV34719). The invention relates to a recombinant or
XX CC associated integrin heterodimer comprising the alpha-10 subunit in
XX CC association with subunit beta (especially beta-1). The heterodimer
XX CC or the subunit alpha-10 can be used as a marker or target of all
XX CC types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
XX CC Cc types can also be used as active ingredients in pharmaceutical
XX CC compositions and vaccines.
XX
XX SQ Sequence 13 AA;

```

```

Query Match 0.8%; Score 9; DB 21; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 700 FTASLDEWT 708
   |||||
Db 1 FTASLDEWT 9

```

```

RESULT 57
AAW35852
ID AAW35852 standard; protein; 1065 AA.
XX
XX AAW35852;
XX
XX 27-APR-1998 (first entry)
XX
XX Human CD11 for use in T lymphocyte veto molecule.
XX
XX Human; CD11; T lymphocyte veto molecule; chimeric molecule;
XX targeting polypeptide; suppression; immune response; treatment;
XX autoimmune disease; allergy; immunological disorder;
XX transplant rejection.
XX
XX Homo sapiens.
XX
XX WO9737687-A1.
XX
XX 16-OCT-1997.
XX
XX 10-APR-1997; 97WO-US05943.
XX

```

PR 10-APR-1996; 96US-0630172.  
 XX (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.  
 PA  
 XX  
 PI Scaerz UD;  
 DR WPI; 1997-512419/47.  
 XX  
 PT T lymphocyte veto molecule comprising response cell activating  
 PT protein - linked to molecule that targets stimulator cell marker.  
 PT used for selective suppression of immune response, e.g. prevention  
 of graft rejection or treatment of auto-immune disease  
 XX  
 PS Claim 37; Pages 61-65; 309pp; English.  
 XX  
 CC A novel T lymphocyte veto molecule is a chimeric molecule  
 CC comprising a protein, e.g. the present sequence, linked to a  
 CC targeting polypeptide that binds a molecule, which differentiates  
 CC a host cell from a tissue graft cell, or selectively targets a  
 CC stimulator cell involved in the autoimmune response.  
 CC A veto molecule, in which the protein binds a molecule that targets  
 CC stimulator cells, can be used to suppress an immune response and  
 CC therefore treat autoimmune diseases, e.g. systemic lupus  
 CC erythematosus, myasthenia gravis, rheumatoid arthritis, insulin  
 CC dependent diabetes mellitus, multiple sclerosis, coeliac disease,  
 CC autoimmune thyroiditis, Addison's or Grave's diseases and  
 CC rheumatoid carditis, allergies and other immunological disorders.  
 CC Where the protein binds a molecule that differentiates graft and  
 CC host cells, the veto molecule can be used to reduce transplant  
 CC rejection. The veto molecule provides specific regulation of  
 CC particular stimulator cells that can kill graft cells or respond  
 CC to autoantigens, but leave other stimulator cells unaffected, e.g.  
 CC CD4 or CD8 positive cells can be regulated without one affecting  
 CC the other. The veto molecule can be administered locally to  
 CC minimise generalised immunosuppression.  
 XX  
 SQ Sequence 1065 AA;  
 QY  
 Query Match 0.8%; Score 9; DB 18; Length 1065;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 566 DVAVGAPLE 574  
 |||||  
 514 DVAVGAPLE 522

RESULT 58  
 AAW81839  
 ID AAW81839 standard; Protein; 1086 AA.  
 XX  
 AC AAW81839;  
 XX  
 DT 05-FEB-1999 (first entry)  
 DE Human LFA-1 alpha chain protein.  
 XX  
 KW LFA-1; alpha chain; human; anti-inflammatory; soluble.  
 XX  
 OS Homo sapiens.  
 XX  
 PN JP10295375-A.  
 XX  
 PD 10-NOV-1998.  
 XX  
 PF 25-APR-1997; 97JP-0109752.  
 XX  
 PR 25-APR-1997; 97JP-0109752.  
 XX  
 PA (DAUC) DAICHI PHARM CO LTD.  
 XX  
 DR WPI; 1999-038275/04.  
 DR N-PSDB; AAV64855.

XX  
 PT Nucleic acid encoding soluble LFA-1 protein - used to identify  
 PT anti-inflammatory agents for use with LFA-1 mediated activity  
 XX  
 PS Disclosure; Page 13; 14pp; Japanese.  
 XX  
 CC This sequence represents a soluble LFA-1 alpha chain protein. The  
 CC encoding nucleic acids can be used for the evaluation of an  
 CC anti-inflammatory agent by a test for the combination between  
 CC ICAM protein and a soluble LFA-1 hetero dimer protein. The soluble  
 CC LFA-1 protein does not associate in the absence of a surfactant.  
 XX  
 SQ Sequence 1086 AA;  
 QY  
 Query Match 0.8%; Score 9; DB 20; Length 1086;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 566 DVAVGAPLE 574  
 |||||  
 538 DVAVGAPLE 546

RESULT 59  
 AAB03973  
 ID AAB03973 standard; Protein; 1145 AA.  
 XX  
 AC AAB03973;  
 XX  
 DT 26-FEB-2001 (first entry)  
 DE  
 KW Alpha\_L chain I domain of leukocyte function associated antigen.  
 KW Leukocyte function associated antigen; LFA-1; I domain; inhibitor;  
 KW ligand; binding; pathology; treatment; screening; detection;  
 KW inflammatory disease; autoimmune disease; reperfusion injury;  
 KW myocardial infarction; stroke; organ transplant;  
 KW respiratory distress syndrome; glomerulonephritis; Crohn's disease;  
 KW T cell mediated disease; rheumatoid arthritis; osteoarthritis;  
 KW spondylitis; thyroid associated ophthalmopathy; Behcet disease;  
 KW sepsis; asthma; chronic bronchitis; silicosis; pulmonary sarcoidosis;  
 KW fibrosis; cystic fibrosis; keloid formation; scar formation;  
 KW atherosclerosis; transplant rejection; lupus;  
 KW inflammatory bowel disease; ulcerative colitis; leukaemia;  
 KW atopic dermatitis; psoriasis; urticaria; uveitis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200060355-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 03-APR-2000; 2000WO-US08841.  
 XX  
 PR 02-APR-1999; 99US-0285477.  
 XX  
 PA (ICOS-) ICOS CORP.  
 PA (ABBO) ABBOTT LAB.  
 XX  
 PI Staunton D, Van Der Vieren M, Hutch J, Fowler K, Orme M,  
 PI Olejniczak ET;  
 DR WPI; 2000-665028/64.  
 DR N-PSDB; AAA54333.  
 XX  
 PT Identifying inhibitor of leukocyte function-associated antigen binding  
 PT to natural ligand for ameliorating inflammatory diseases, by measuring  
 PT antigen-ligand binding in presence and absence of test compound  
 XX  
 PS Example 2; Page 51-54; 66pp; English.  
 XX  
 CC Inhibitors of leukocyte function-associated antigen binding are  
 CC useful for manufacturing a medicament for ameliorating a pathology

CC arising from LFA-1 binding to its ligand such as inflammatory  
 CC diseases, autoimmune diseases, reperfusion injury, myocardial  
 CC infarction, stroke, organ transplant, adult respiratory distress  
 CC syndrome, acute glomerulonephritis, Crohn's disease, T cell mediated  
 CC diseases, rheumatoid arthritis, osteoarthritis, spondylitis, thyroid  
 CC associated ophthalmopathy, Behcet disease, sepsis, asthma, chronic  
 CC bronchitis, allergic respiratory distress syndrome, chronic pulmonary  
 CC inflammatory disease, such as chronic obstructive pulmonary disease,  
 CC silicosis, pulmonary sarcoidosis, fibrosis, cystic fibrosis, keloid  
 CC formation, scar formation, atherosclerosis, transplant rejection  
 CC disorders, such as graft versus host reaction and allograft  
 CC rejection, chronic glomerulonephritis, lupus, inflammatory bowel  
 CC disease, such as ulcerative colitis, proliferative lymphocyte  
 CC diseases, such as leukemia, and inflammatory dermatoses, such as  
 CC atopic dermatitis, psoriasis, urticaria, and uveitis.  
 XX  
 SO Sequence 1145 AA:  
 Query Match 0.8%; Score 9; DB 21; Length 1145;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 566 DVAVGAPLE 574  
 DB 513 DVAVGAPLE 521  
 RESULT 60  
 AAR05782  
 ID AAR05782 standard; Protein; 1170 AA.  
 XX  
 AC AAR05782;  
 XX  
 DT 23-AUG-1990 (first entry)  
 XX  
 DE LFA-1 alpha subunit.  
 XX  
 KM Lymphocyte function associated antigen; inflammation; metastasis.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..25  
 FT Domain /label= sig\_peptide  
 FT Domain /label= extracellular\_domain  
 FT Domain /label= transmembrane\_domain  
 FT Region 1118..1170  
 FT Peptide /label= cytoplasmic\_tail  
 FT Peptide 118..132  
 FT Peptide /label= tryptic\_fragment  
 FT Peptide 226..237  
 FT Peptide /label= tryptic\_fragment  
 FT Peptide 282..288  
 FT Peptide /label= tryptic\_fragment  
 FT Peptide 433..441  
 FT Peptide /label= tryptic\_fragment  
 FT Peptide 521..531  
 FT Peptide /label= tryptic\_fragment  
 FT Peptide 569..582  
 FT Peptide /label= tryptic\_fragment  
 FT Peptide 590..604  
 FT Peptide /label= tryptic\_fragment  
 FT Peptide 831..844  
 FT Peptide /label= tryptic\_fragment  
 FT Peptide 957..974  
 FT Peptide /label= tryptic\_fragment  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 89..91  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 188..190  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 649..651

FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 670..672  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 726..728  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 730..732  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 862..864  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 885..887  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 897..899  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 1060..1062  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 1071..1073  
 FT Modified-site /label= N-glycosylation\_site  
 FT Modified-site 1140  
 FT Modified-site /label= serine\_phosphorylation\_site  
 FT Modified-site 1145  
 FT Modified-site /label= serine\_phosphorylation\_site  
 FT Modified-site 1163  
 FT Modified-site /label= serine\_phosphorylation\_site  
 FT Modified-site 1165  
 FT Modified-site /label= serine\_phosphorylation\_site  
 PN EP362526-A.  
 PD 11-APR-1990.  
 XX  
 XX 17-AUG-1989; 89EP-0115160.  
 XX  
 PR 23-AUG-1988; 88US-0235227.  
 PR 09-MAR-1989; 89US-0321017.  
 XX  
 PA (DANA-) DANA FARBER CANCER.  
 XX  
 PI Springer TA, Larson R;  
 PI  
 DR WPI; 1990-108985/15.  
 DR N-PSDB; AAO03842.  
 XX  
 PT Pure alpha subunit of lymphocyte function associated antigen -  
 PT and encoding DNA sequences, useful eg for suppressing  
 PT inflammation or metastasis  
 XX  
 PS Disclosure; Fig 3; 27pp; English.  
 XX  
 CC The alpha-subunit (a-SU) can bind to ICM-1 (or other natural  
 CC ligands) on the surface of cells, and can associate with the beta-SU  
 CC to form a heterodimer (also able to bind to ICM-1). a-SU, and its  
 CC derivs., are useful in suppressing inflammation, metastasis and its  
 CC growth of a-SU expressing tumour cells and is used in the treatment  
 CC of viral infections.  
 CC The pref. dose is 1 pg - 10 mg/kg.  
 XX  
 SO Sequence 1170 AA:  
 Query Match 0.8%; Score 9; DB 11; Length 1170;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 566 DVAVGAPLE 574  
 DB 538 DVAVGAPLE 546  
 RESULT 61  
 AAR80107  
 ID AAR80107 standard; Protein; 1170 AA.  
 XX  
 AC AAR80107;  
 XX

DT 04-MAR-1996 (first entry)  
 XX LFA-1 alpha subunit CD11a.  
 DE CD11a; leucocyte function-associated antigen-1; LFA-1; integrin.  
 XX  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Peptide 1..25 /label= Sig\_peptide  
 FT 32..79 /label= Repeat  
 FT /note= "Repeat I"  
 FT 82..132 /label= Repeat  
 FT /note= "Repeat II"  
 FT 339..391 /label= Repeat  
 FT /note= "Repeat III"  
 FT 392..446 /label= Repeat  
 FT /note= "Repeat IV"  
 FT 447..508 /label= Repeat  
 FT /note= "Repeat V"  
 FT 509..567 /label= Repeat  
 FT /note= "Repeat VI"  
 FT 568..629 /label= Repeat  
 FT /note= "Repeat VII"  
 FT 170..349 /label= IDomain  
 FT /note= "internal domain"  
 FT 1089..1112 /label= Trans  
 FT /note= "transmembrane domain"  
 FT 1113..1170 /label= Cyto  
 FT /note= "cytoplasmic domain"  
 FT  
 PN WO9528170-A1.  
 XX  
 PD 26-OCT-1995.  
 XX  
 PF 19-APR-1995; 95WO-US04886.  
 XX  
 PR 19-APR-1994; 94US-0229513.  
 XX  
 PA (UNIV ) UNIV KANSAS.  
 PI  
 XX Benedict S, Chan MA, Siahaan TJ, Tibbets SA;  
 DR WPI; 1995-373631/48.  
 XX  
 PT Changing peptide reactivity via conjugation with a second peptide  
 PT causes change in conformation of first peptide, pref. LFA-1 and  
 PT ICM-1 functional domain derived peptide(s)  
 PT  
 XX  
 PS Claim 11; Page 34-37; 64pp; English.  
 XX  
 CC Functional domains derived from LFA-1 alpha subunit CD11a (AAR80107),  
 CC LFA-1 beta subunit CD18 (AAR80107) and ICM-1 CD43 (AAR80110) are used  
 CC as the basis of short, synthetic peptides (AAR80109, AAR80111-24) that  
 CC modulate ICM/LFA binding interaction. CD11a is obcd. from human  
 CC PMA-stimulated T-lymphocyte HL-60 cells.  
 CC  
 SO Sequence 1170 AA;

QY 566 DVAVGAPLE 574  
 |||||  
 Db 538 DVAVGAPLE 546  
 RESULT 62  
 ABB76377  
 ID ABB76377 standard; Protein; 1170 AA.  
 XX  
 AC ABB76377;  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX  
 DE Lymphocyte function associated antigen-1.  
 XX  
 KW Lymphocyte function associated antigen-1; LFA-1; human;  
 KW alpha/beta protein; ligand binding.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Domain 153..332 /label= I\_domain  
 FT  
 PN WO200231511-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PF 12-OCT-2001; 2001WO-US32047.  
 XX  
 PR 12-OCT-2000; 2000US-239750P.  
 XX  
 PA (ICOS-) ICOS CORP.  
 XX  
 PI Stauton DE;  
 XX  
 DR WPI; 2002-471361/50.  
 XX  
 PT Modulating binding interactions between alpha/beta proteins comprising  
 PT allosteric regulatory sites, and a binding partner, by contacting with  
 PT an allosteric effector molecule that interacts with the regulatory site  
 PT  
 PS Disclosure; Fig 2; 163pp; English.  
 XX  
 CC The present sequence is the protein sequence of lymphocyte function  
 CC associated antigen-1 (LFA-1), which includes an integrin (I)  
 CC domain. The invention provides methods of modulating the binding  
 CC interaction of a first molecule, which is not LFA-1 or an I  
 CC domain-containing fragment of LFA-1, and a binding partner. The  
 CC first molecule has an alpha/beta domain structure comprising an  
 CC allosteric regulatory site. Modulation of the binding interaction  
 CC involves contacting with an allosteric effector molecule that  
 CC interacts with the regulatory site. In vivo methods are expected  
 CC to alleviate and/or prevent pathological states arising from  
 CC aberrant binding activity. Methods for identifying modulators are  
 CC also provided.  
 CC  
 SO Sequence 1170 AA;

Query Match 0.8%; Score 9; DB 23; Length 1170;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574  
 |||||  
 Db 538 DVAVGAPLE 546

RESULT 63  
 AA080251  
 ID AA080251 standard; Protein; 1170 AA.

```

XX AC AAU80251;
XX DT 15-JUL-2002 (first entry)
XX DE Human integrin 1 alpha-L subunit protein.
XX KW Integrin; antinflammatory; immunosuppression; nephritis; dermatitis;
XX KW inflammatory disease; autoimmune disorder; Crohn's disease;
XX KW human immunodeficiency virus; HIV; myocardial infarction;
XX KW Sjogren's syndrome; rheumatoid arthritis.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 214 /note="Encoded by TGC"
XX FT WO200218583-A2.
XX PN 07-MAR-2002.
XX PD 31-AUG-2001; 2001WO-US27227.
XX PF 01-SEP-2000; 2000US-229700P.
XX PR (BLOO-) CENT BLOOD RES INC.
XX PA Springer TA, Shimoaka M, Lu C;
XX PI WPI; 2002-382964/41.
XX DR N-PSDB; ABR50045.
XX XX
XX PT Modified integrin-I or integrin I-like domain polypeptide useful as an
XX PT immunogen to produce antibodies specific to polypeptide, comprises a
XX PT disulfide bond such that polypeptide is stabilized in a desired
XX PT conformation
XX PS Disclosure; Page 105-108; 112pp; English.
XX XX
XX CC This invention relates to a modified integrin-I or integrin I-like
XX CC domain polypeptide comprising at least one disulfide bond so that the
XX CC domain is stabilized in a desired conformation. The polypeptide of
XX CC the invention may have antinflammatory or immunosuppressive activities.
XX CC The polypeptides of the invention have an open conformation and are
XX CC useful as immunogens to produce antibodies that selectively bind to
XX CC integrin I-domain, and for identifying a modulator of integrin activity,
XX CC or of interaction of an integrin and a cognate ligand. The polypeptide
XX CC of the invention, or antibodies (preferably anti-LFA-1 antibody)
XX CC is useful for treating or preventing an integrin mediated disorder which
XX CC is an inflammatory or autoimmune disorder in a subject and for
XX CC inhibiting the binding of an integrin to a cognate ligand in a subject.
XX CC A therapeutic composition comprising the peptide of the invention is
XX CC useful for treating an integrin mediated disorder such as Crohn's
XX CC disease, nephritis; human immunodeficiency virus (HIV), myocardial
XX CC infarction, Sjogren's syndrome, rheumatoid arthritis, dermatitis. The
XX CC polypeptides and/or active or antigenic fragments are useful as
XX CC reagents for diagnosis of integrin-mediated disorders. The present
XX CC sequence represents the human integrin-1 alpha-L protein subunit used to
XX CC generate the mutant polypeptides of the invention.
XX SQ Sequence 1170 AA;

```

```

Query Match 0.8%; Score 9; DB 23; Length 1170;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 566 DVAVGAPLE 574
   |||||
DB 538 DVAVGAPLE 546

```

RESULT 64

```

AAV32248
ID AAV32248 standard; peptide; 8 AA.
XX AC AAV32248;
XX DT 15-FEB-2000 (first entry)
XX DE Bovine integrin subunit alpha-10 tryptic peptide 4.
XX KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
XX KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;
XX KW inflammation; therapy; cartilage; chondrocyte; osteoblast;
XX KW fibroblast; vaccine; marker.
XX OS Bos taurus.
XX FH WO9951639-A1.
XX PN 14-OCT-1999.
XX PD 31-MAR-1999; 99WO-SE00544.
XX PF 02-APR-1998; 98SE-0001164.
XX PR 28-JAN-1999; 99SE-0000319.
XX PA (ACTI-) ACTIVE BIOTECH AB.
XX PI Lundgren-Akerlund E;
XX DR WPI; 2000-052639/04.
XX XX
XX PT New isolated integrin subunit alpha-10, used as a marker or target
XX PT molecule for cells during development, regeneration and pathological
XX PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
XX PT inflammation
XX PS Example 2; Fig 2; 90pp; English.
XX XX
XX CC This sequence represents tryptic peptide 4 of integrin subunit
XX CC alpha-10 (ISa10) isolated from bovine chondrocytes. 6 Tryptic
XX CC peptides (see AAV32245-50) were obtained. PCR primers (see
XX CC AA234721-24) based on peptides 1 and 2 were used to obtain bovine
XX CC ISa10 cDNA fragments. These were used to screen a human
XX CC articular chondrocyte cDNA library in order to isolate human ISa10
XX CC cDNA (see AA234719). The invention relates to a recombinant or
XX CC isolated integrin heterodimer comprising the alpha-10 subunit in
XX CC association with subunit beta (especially beta-1). The heterodimer
XX CC or the subunit alpha-10 can be used as a marker or target of all
XX CC types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
XX CC They can also be used as active ingredients in pharmaceutical
XX CC compositions and vaccines.
XX SQ Sequence 8 AA;

```

```

Query Match 0.7%; Score 8; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 551 FAMGALPD 538
   |||||
DB 1 FAMGALPD 8

```

```

RESULT 65
AAW99005
ID AAW99005 standard; peptide; 17 AA.
XX AC AAW99005;
XX DT 11-MAY-1999 (first entry)
XX DE I domain peptide P10.

```

RESULT 65

KW Cyclic integrin binding peptide; integrin alpha-2i domain; inhibitor;  
 KW collagen I; collagen IV; laminin-1; cell migration; cancer;  
 KW cardiovascular disease; periodontal disease.  
 OS Synthetic.  
 XX  
 XX WO902551-A1.  
 XX  
 XX 21-JAN-1999.  
 XX  
 XX PD  
 XX PF 09-JUL-1998; 98WO-F100579.  
 XX  
 XX PR 11-JUL-1997; 97US-0893526.  
 XX  
 XX PA (HEIN/) HEINO J.  
 XX PA (IVAS/) IVASKA J.  
 XX PA (KABE/) KAEPYLAE J.  
 XX  
 XX PI Heino J, Ivaska J, Kaepylae J;  
 XX  
 XX DR WPI; 1999-120775/10.  
 XX  
 XX PT Cyclic integrin binding peptides - used to inhibit  
 XX PT integrin-dependent cell migration  
 XX  
 XX PS Example 3; Page 25; 59pp; English.  
 XX  
 CC The present sequence invention describes cyclic integrin binding peptides  
 CC comprising the amino acid sequence RKK, preferably RKKI. Also described  
 CC in the present invention are: (1) a cyclic peptide comprising the amino  
 CC acid sequence X1RKKHX2n where X is any amino acid and n=1-4; (2) a  
 CC cyclic integrin binding peptide comprising the amino acid sequence  
 CC CTRKKHMC or CTRKKHNAQC; (3) a pharmaceutical composition comprising  
 CC one of the above integrin binding peptides; and (4) a binding assay for  
 CC identifying integrin binding agents, comprising: (a) biotinylating the  
 CC integrin binding agent to be assayed; (b) reacting the biotinylated  
 CC agent with an immobilised recombinant alpha 2i domain or domain-derived  
 CC peptides in conditions suitable for binding; (c) washing the solid  
 CC support carrying the bound agent; (d) adding a labeled biotin-binding  
 CC agent; and (e) detecting any bound integrin binding agent. The integrin  
 CC binding peptides can be used for inhibiting integrins. They can also be  
 CC used for inhibiting integrin-dependent cell migration, such as  
 CC associated with cancer, cardiovascular disease or a periodontitis  
 CC condition. They can also be used for inhibiting the adhesion of  
 CC platelets to collagen or collagen induced platelet aggregation in a  
 CC patient. The present sequence represents a peptide from the present  
 CC invention.  
 CC  
 CC SQ Sequence 17 AA;  
 CC  
 CC Query Match 0.7%; Score 8; DB 20; Length 17;  
 CC Best Local Similarity 100.0%; Pred. No. 7.1;  
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 275 TDGESHG 282  
 CC Db 1 TDGESHG 8  
 CC  
 CC RESULT 66  
 CC AAY12862  
 CC ID AAY12862 standard; Protein; 47 AA.  
 CC XX  
 CC AC AAY12862;  
 CC XX  
 CC DT 21-JUN-1999 (first entry)  
 CC XX  
 CC DE Human 5' EST secreted protein SEQ ID NO:452.  
 CC XX  
 CC XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 CC KW forensic; gene therapy; chromosome mapping; signal peptide;  
 CC KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 CC KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 CC KW

KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO906549-A2.  
 XX  
 XX PN 11-FEB-1999.  
 XX  
 XX PD  
 XX PF 31-JUL-1998; 98WO-1B01231.  
 XX  
 XX PR 01-AUG-1997; 97US-0905279.  
 XX  
 XX PA (GEST ) GENSET.  
 XX  
 XX PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 XX  
 XX DR WPI; 1999-153779/13.  
 XX DR N-PSDB; AAX51640.  
 XX  
 XX PT New nucleic acids encoding human secreted proteins - obtained from  
 XX PT CDNA libraries derived from testis, ovary, uterus and spleen tissue  
 XX  
 XX PS Claim 34; Page 490; 522pp; English.  
 XX  
 CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12681 to  
 CC AAY12913, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 CC  
 CC SQ Sequence 47 AA;  
 CC  
 CC Query Match 0.7%; Score 8; DB 20; Length 47;  
 CC Best Local Similarity 100.0%; Pred. No. 17;  
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1133 GLLLLALL 1140  
 CC Db 12 GLLLLALL 19  
 CC  
 CC RESULT 67  
 CC AAG00326  
 CC ID AAG00326 standard; Protein; 66 AA.  
 CC XX  
 CC AC AAG00326;  
 CC XX  
 CC DT 06-OCT-2000 (first entry)  
 CC XX  
 CC DE Human secreted protein, SEQ ID NO: 4407.  
 CC XX  
 CC XX Human; 5' EST; expressed sequence tag; secreted protein; CDNA isolation;  
 CC KW gene therapy; chromosome mapping.  
 CC XX  
 CC OS Homo sapiens.  
 CC XX  
 CC PN EP1033401-A2.  
 CC XX  
 CC PD 06-SEP-2000.  
 CC XX



```

PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
PA (GETH ) GENENTECH INC.
XX
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
DR WPI, 1999-551358/46.
DR N-PSDB; AAZ33963.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
XX PS Claim 12; Fig 33; 530pp; English.
XX
XX CC The present invention describes secreted and transmembrane polypeptides
XX and their polynucleotides. The nucleotide sequences are useful as
XX CC sources of probes, primers, for chromosome mapping, and for generation
XX of antisense sequences. They can also be used to create transgenic
XX animals. The proteins can be used to treat a variety of diseases and
XX disorders, depending on their function. Diseases that may be treated
XX include blood coagulation disorders, cancers and cellular adhesion
XX disorders. They may also be used to raise antibodies. AAZ33891 to
XX AA23338, and AA41685 to AA41774 represent polynucleotide and
XX CC polypeptide sequence given in the exemplification of the present
XX invention.
XX
SQ Sequence 67 AA;
Query Match 0.7%; Score 8; DB:20; Length 67;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1133 GLLLLALL 1140
12 GLLLLALL 19
Db
RESULT 69
AAB44252
ID AAB44252 standard; Protein; 67 AA.
AC AAB44252;
XX
XX DT 08-FEB-2001 (first entry)
XX
DE Human PRO617 (UNQ353) protein sequence SEQ ID NO:85.
XX
XX KW Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;
XX expressed sequence tag; detection; cancer.
XX
XX OS Homo sapiens.
XX
XX PN MO200053756-A2.
XX
XX PD 14-SEP-2000.
XX
XX PF 18-FEB-2000; 2000MO-US04341.
XX
XX PR 08-MAR-1999; 99MO-US05028.
XX PR 12-MAR-1999; 99US-0123957.
XX PR 29-MAR-1999; 99US-0126773.
XX PR 21-APR-1999; 99US-0130232.
XX PR 28-APR-1999; 99US-0131445.
XX PR 14-MAY-1999; 99US-0134287.
XX PR 23-JUN-1999; 99US-0141037.
XX PR 26-JUL-1999; 99US-0145698.
XX PR 29-OCT-1999; 99US-0162506.

```

```

PR 30-NOV-1999; 99MO-US28313.
PR 02-DEC-1999; 99MO-US28551.
PR 02-DEC-1999; 99MO-US28565.
PR 16-DEC-1999; 99MO-US30095.
PR 30-DEC-1999; 99MO-US31243.
PR 30-DEC-1999; 99MO-US31274.
PR 05-JAN-2000; 2000MO-US00219.
PR 06-JAN-2000; 2000MO-US00277.
PR 06-JAN-2000; 2000MO-US00376.
XX
XX PA (GETH ) GENENTECH INC.
XX
XX PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL,
XX PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gertschen ME,
XX PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan RJ,
XX PI Kiyavits IJ, Kuo SS, Napier MA, Pan J, Paoletti NF, Roy MA,
XX PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX DR WPI; 2000-611443/58.
XX DR N-PSDB; AAC78478.
XX
XX PT Novel PRO polypeptides and polynucleotides used in detection methods,
XX PT to target bioactive molecules to specific cells, and to modulate
XX PT cellular activities -
XX
XX PS Claim 12; Fig 33; 636pp; English.
XX
XX CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
XX sequence tag) sequences which encode secreted or transmembrane PRO
XX polypeptides. The PRO polynucleotides and polypeptides have cytostatic
XX activity. The polynucleotides and polypeptides can be used for detecting
XX CC the presence of PRO polypeptides in samples, for linking bioactive
XX CC molecules to cells and for modulating biological activities of cells,
XX using the polypeptides for specific targeting. The polypeptide targeting
XX CC can be used to kill the target cells, e.g. for the treatment of cancers.
XX CC The polypeptide pairs provide specific targeting of bioactive molecules
XX CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
XX CC the isolation of the PRO polynucleotide sequences.
XX
SQ Sequence 67 AA;
Query Match 0.7%; Score 8; DB:21; Length 67;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1133 GLLLLALL 1140
12 GLLLLALL 19
Db
RESULT 70
AAB19581
ID AAB19581 standard; Protein; 67 AA.
AC AAB19581;
XX
XX DT 22-JAN-2001 (first entry)
XX
DE Human PRO617 used to treat ocular disease.
XX
XX KW PRO617; human; ocular disease; retinopathy; maculopathy; therapy;
XX retinitis pigmentosa; macular degeneration; retinal detachment;
XX KW retinal tear; macular hole; myopia; traumatic chorioretinopathy;
XX acute retinal necrosis syndrome; confusion; edema;
XX KW retinal vision occlusion; vascular disease; retinal vasculitis;
XX thrombocytopenic purpura; uveitis; retinal occlusion.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 15..27
XX FT Protein /Label= Signal_peptide 28..67

```



FT 1  
FT Modified-site 41.43 /label= Mature\_Pro617  
FT /note= "O-phosphorylated"  
XX  
XX MO200053760-A2.  
XX  
XX 14-SEP-2000.  
XX  
XX 10-MAR-2000; 2000WO-US06319.  
XX  
XX 12-MAR-1999; 99US-0123957.  
XX  
XX (GENENTECH INC.  
XX  
XX Ferrara N, Goddard A, Gurney AL, Hebert C, Henzel WJ, Kabakoff RC;  
XX Klein RD, Kljavin ID, Kuo SS, La Fleur M, Wood WI,  
XX MPI: 2000-587437/55.  
XX N-PSDB; AAA88518.  
XX  
XX Novel PRO polypeptides useful for preventing or rescuing retinal cells  
XX from injury caused by ocular diseases such as retinitis pigmentosa,  
XX retinopathy, retinal degenerative diseases, degenerative myopia,  
XX uveitis -  
XX  
XX Claim 2; Fig 10; 140pp; English.  
XX  
XX The present sequence is that of human PRO617, as predicted from a  
XX cDNA clone (see AAA88518) isolated from a foetal kidney tissue cDNA  
XX library using probes and primers (see AAA88535-37) based on a cDNA  
XX clone isolated from an expressed sequence tag database screening  
XX using about 950 secreted protein extracellular domains. PRO617  
XX has a predicted mol.wt. of 6,981 and a pI of about 7.47. A  
XX method for producing PRO polypeptides, including PRO617, using a  
XX host cell transformed with a vector comprising a PRO nucleic acid is  
XX claimed. The invention relates to the use of PRO polypeptides to delay,  
XX prevent or rescue retinal cells such as retinal neurons selected from  
XX photoreceptors, retinal ganglion cells, displaced retinal ganglion  
XX cells, amacrine cells, displaced amacrine cells, horizontal and  
XX bipolar neurons, and supportive cells (including Muller cells and  
XX pigment epithelial cells) from injury and degradation. The retinal  
XX cells are preferably photoreceptors and photoreceptor cell injury or  
XX death is caused by retinal injury, light or environmental trauma or  
XX by an ocular disease selected from retinitis pigmentosa, macular  
XX degeneration, including age-related, retinal detachment, retinal  
XX tears, retinopathy, retinal degenerative diseases, macular holes,  
XX degenerative myopia, acute retinal necrosis syndrome, traumatic  
XX chorioretinopathies or contusion such as Pott's retinopathy,  
XX edema, ischemic conditions such as central or branch retinal vision  
XX occlusion, collagen vascular diseases, thrombocytopenic purpura,  
XX uveitis, retinal vasculitis and occlusion associated with Eales  
XX disease and systemic lupus erythematosus (claimed).  
XX  
XX Sequence 67 AA:  
SQ  
Query Match 0.7%; Score 8; DB 21; Length 67;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1133 GLLALL 1140  
DB 12 GLLALL 19  
RESULT 71  
AAU28155  
ID AAU28155 standard; Protein; 67 AA.  
XX  
XX AAU28155;  
AC  
XX 18-DEC-2001 (first entry)  
XX  
XX Novel human secretory protein, Seq ID No 324.  
DE

XX  
XX Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;  
XX ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
XX transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
XX amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
XX ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
XX gut protection; lung; liver fibrosis; immune deficiency; infection;  
XX severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
XX multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
XX fertility; analgesic; pain; antigen.  
XX  
XX Homo sapiens.  
XX  
XX WO200166689-A2.  
XX  
XX 13-SEP-2001.  
XX  
XX 05-MAR-2001; 2001WO-US04942.  
XX  
XX 07-MAR-2000; 2000US-0519705.  
XX 19-MAY-2000; 2000US-0574454.  
XX 17-JUN-2000; 2000US-0596193.  
XX 14-JUL-2000; 2000US-0616847.  
XX 19-SEP-2000; 2000US-0665363.  
XX 20-OCT-2000; 2000US-0693267.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P,  
XX Zhao Q, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AD, Wang J;  
XX MPI: 2001-589934/66.  
XX N-PSDB; AAS45055.  
XX  
XX Novel polypeptides and nucleic acids obtained from cDNA libraries  
XX prepared from various human tissues, for diagnosis and treatment of  
XX cancer, neurological, inflammatory, and autoimmune disorders -  
XX  
XX Example 4; SEQ ID No 324; 107pp; English.  
XX  
XX The invention relates to novel isolated human secreted polypeptides (I)  
XX and polynucleotides (II). (I) and (II) are useful for treating  
XX inflammatory conditions such as arthritis, nephritis, Crohn's disease,  
XX ischaemia-reperfusion injury, shock, sepsis, immune responses, and is  
XX involved in increasing haematopoiesis, stem cell survival, bone growth  
XX and remodeling. (I), (II) and modulators of (II) are useful for  
XX prophylaxis or treatment of one or more cancers. (II) is also useful for  
XX creating transgenic animals useful for studying the in vivo activities of  
XX the polypeptide as well as for studying modulators of the polypeptides.  
XX (I) induces the proliferation of neural cells and regeneration of nerve  
XX and brain tissue and is useful for the treatment of central and  
XX peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
XX Parkinson's disease, Huntington's disease, and amyotrophic lateral  
XX sclerosis. In addition, (I) is involved in chemotactic or chemokinetic  
XX activity, regulation of haematopoiesis and is useful for treating myeloid  
XX or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
XX and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
XX tissue growth, and in tissue repair, healing of burns, incisions,  
XX ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
XX disorders, or periodontal disease. Furthermore, (I) is also useful for  
XX gut protection or regeneration and treatment of lung or liver fibrosis,  
XX reperfusion injury in various tissues, various immune deficiencies and  
XX disorders including severe combined immunodeficiency (SCID), bacterial or  
XX fungal infections, autoimmune disorders e.g. multiple sclerosis,  
XX rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
XX reactions and conditions, such as asthma or other respiratory problems.  
XX In addition, (I) affects biorythms or circadian cycles of rhythms,  
XX fertility, metabolism, catabolism, anabolism, storage or elimination of  
XX dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
XX analgesic effects or other pain reducing effects, immunoglobulin like  
XX activity and can act as an antigen in a vaccine composition to raise an  
XX immune response. AAU28020-AAU28395 represent novel human secreted protein  
XX amino acid sequences of the invention.

XX Sequence 67 AA; Score 8; DB 22; Length 67;  
SQ Query Match Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1133 GLLLLALL 1140  
|||  
12 GLLLLALL 19

Db

RESULT 72  
AAU74938  
ID AAU74938 standard; Protein; 67 AA.  
XX AAU74938;  
XX DT 23-APR-2002 (first entry)  
XX DE Human clone DNA48309-1280 amino acid sequence of PRO617 protein.  
XX KW Human; clone DNA48309-1280; PRO617; immune response modulator;  
XX KW PRO polypeptide; immune related disorder; systemic lupus erythematosus;  
XX KW rheumatoid arthritis; osteoarthritis; systemic sclerosis; sarcoidosis;  
XX KW autoimmune thrombocytopenia; thyroiditis; diabetes mellitus; anaemia;  
XX KW multiple sclerosis; inflammatory bowel disease; allergic rhinitis.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX FT Peptide 15..27  
XX FT Protein /label= Signal\_peptide  
XX FT 28..67  
XX FT /label= Mature\_PRO617\_protein  
XX PN WO200192331-A2.  
XX PD 06-DEC-2001.  
XX PF 25-MAY-2001; 2001WO-US17092.  
XX PR 30-MAY-2000; 2000WO-US14941.  
XX PR 02-JUN-2000; 2000WO-US15264.  
XX PR 01-DEC-2000; 2000WO-US32678.  
XX PA (GENTH ) GENENTECH INC.  
XX PI Eaton DL, Fong S, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL;  
XX PI Watanabe CK, Wood WI, Zhang Z;  
XX DR MPI: 2002-075461/10.  
XX DR N-PSDB; ABK14010.  
XX PT Isolated PRO221, PRO617, PRO1030, PRO4302 polypeptides, useful for  
XX PT treating immune disorders such as thyroiditis, diabetes mellitus,  
XX PT allergic disease, asthma, allergic rhinitis, atopic dermatitis -  
XX PS Claim 10; Fig 4; 104dp; English.  
XX  
XX The present invention relates to a new polypeptide having at least 80%  
XX amino acid sequence identity to fully defined amino acid sequence of  
XX PRO221 (AAU74937), PRO617 (AAU74938), PRO1030 (AAU74939) or PRO4302  
XX (AAU74940) as given in specification. The PRO polypeptides of the  
XX invention are useful for treating an immune related disorder such as  
XX systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,  
XX CC systemic sclerosis, sarcoidosis, autoimmune thrombocytopenia,  
XX CC thyroiditis, diabetes mellitus, multiple sclerosis, inflammatory bowel  
XX CC disease, allergic rhinitis and anaemia. The methods of the invention  
XX are useful for identifying an antisense nucleic acid and for enhancing  
XX the infiltration of inflammatory cells into the tissue of a mammal. The  
XX anti-PRO antibodies are useful for determining the presence of PRO  
XX polypeptides and for diagnosing an immune-related disease in a mammal.

CC The molecules of the invention are useful for identifying a compound  
CC that inhibits the activity of PRO polypeptides or expression of a gene  
CC encoding the PRO polypeptide. PRO polypeptides are also useful for  
CC increasing or decreasing the infiltration of inflammatory cells from the  
CC vasculature into a tissue of a mammal, increasing or decreasing activity  
CC or proliferation of T-lymphocytes. The present amino acid sequence  
CC represents the PRO617 polypeptide of the invention. The PRO617 sequence  
CC was isolated from the clone DNA48309-1280.  
XX  
SQ Sequence 67 AA; Score 8; DB 23; Length 67;  
Query Match Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1133 GLLLLALL 1140  
|||  
12 GLLLLALL 19

Db

RESULT 73  
ABB17942  
ID ABB17942 standard; Protein; 69 AA.  
XX ABB17942;  
XX DT 23-JAN-2002 (first entry)  
XX DE Human nervous system related polypeptide SEQ ID NO 6599.  
XX KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
XX KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;  
XX KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;  
XX KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
XX KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
XX KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
XX KW neurological disease; infection; nephrotropic; gene therapy; vaccine.  
XX OS Homo sapiens.  
XX PN WO200159063-A2.  
XX PD 16-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01334.  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.  
XX PR 17-MAR-2000; 2000US-0190076.  
XX PR 18-APR-2000; 2000US-0198123.  
XX PR 19-MAY-2000; 2000US-0205515.  
XX PR 07-JUN-2000; 2000US-0209467.  
XX PR 28-JUN-2000; 2000US-0214886.  
XX PR 30-JUN-2000; 2000US-0215135.  
XX PR 07-JUL-2000; 2000US-0216647.  
XX PR 07-JUL-2000; 2000US-0216880.  
XX PR 11-JUL-2000; 2000US-0217487.  
XX PR 11-JUL-2000; 2000US-0217496.  
XX PR 14-JUL-2000; 2000US-0218290.  
XX PR 26-JUL-2000; 2000US-0220963.  
XX PR 26-JUL-2000; 2000US-0220964.  
XX PR 14-AUG-2000; 2000US-0224518.  
XX PR 14-AUG-2000; 2000US-0224519.  
XX PR 14-AUG-2000; 2000US-0225213.  
XX PR 14-AUG-2000; 2000US-0225214.  
XX PR 14-AUG-2000; 2000US-0225267.  
XX PR 14-AUG-2000; 2000US-0225267.  
XX PR 14-AUG-2000; 2000US-0225268.  
XX PR 14-AUG-2000; 2000US-0225270.  
XX PR 14-AUG-2000; 2000US-0225447.





Query Match 0.7%; Score 8; DB 22; length 76;  
Best Local Similarity 100.0%; Pred.No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1133 GLLTALL 1140  
Db 12 GLLTALL 19

Search completed: July 16, 2003, 08:32:18  
Job time : 52 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 08:13:03 ; Search time 14 Seconds  
(without alignments)  
46.236 Million cell updates/sec

Title: US-09-647-544-7  
Perfect score: 114  
Sequence: 1 KLGFPAHKKIPPEEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 140335

Minimum DB seq length: 0  
Maximum DB seq length: 22

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	37.7	20	1	US-08-214-770-1
2	43	37.7	20	5	PCT-US95-02885-1
3	43	37.7	20	5	PCT-US95-07542-15
4	39	34.2	20	6	5225193-7
5	36	31.6	20	6	5225193-2
6	36	31.6	21	1	US-08-073-028-25
7	36	31.6	21	1	US-08-127-499A-15
8	36	31.6	21	1	US-08-482-847-15
9	36	31.6	21	4	US-08-554-616-25
10	36	31.6	22	1	US-08-073-028-24
11	36	31.6	22	4	US-08-554-616-24
12	35	30.7	21	5	PCT-US95-07542-11
13	34	29.8	13	1	US-08-056-200-11
14	34	29.8	13	1	US-07-987-272A-6
15	34	29.8	13	2	US-08-800-644-11
16	34	29.8	18	2	US-08-649-991-49
17	34	29.8	19	2	US-08-690-011A-4
18	34	29.8	19	4	US-09-299-495F-4
19	34	29.8	20	1	US-08-199-776-21
20	34	29.8	20	3	US-08-663-731-21
21	34	29.8	20	3	US-08-879-338-21
22	34	29.8	20	4	US-09-406-781-1
23	34	29.8	20	4	US-09-181-941-15
24	34	29.8	20	4	US-09-293-238B-21
25	34	29.8	20	5	PCT-US95-02044-21
26	34	29.8	21	1	US-08-199-776-5
27	34	29.8	21	1	US-08-555-579-1

28	34	29.8	21	3	US-08-663-731-5	Sequence 5, Appli
29	34	29.8	21	3	US-08-879-338-5	Sequence 5, Appli
30	34	29.8	21	3	US-08-879-338-26	Sequence 26, Appli
31	34	29.8	21	4	US-09-293-238B-5	Sequence 5, Appli
32	34	29.8	21	4	US-09-293-238B-26	Sequence 26, Appli
33	34	29.8	21	5	PCT-US95-02044-5	Sequence 5, Appli
34	33.5	29.4	22	6	5104631-11	Patent No. 5304631
35	33	28.9	14	1	US-08-477-108A-10	Sequence 10, Appli
36	33	28.9	15	1	US-08-199-776-20	Sequence 20, Appli
37	33	28.9	15	3	US-08-663-731-20	Sequence 20, Appli
38	33	28.9	15	3	US-08-879-338-20	Sequence 20, Appli
39	33	28.9	15	4	US-09-293-238B-20	Sequence 20, Appli
40	33	28.9	15	5	PCT-US95-02044-20	Sequence 20, Appli
41	33	28.9	21	4	US-08-584-043A-81	Sequence 81, Appli
42	32	28.1	16	4	US-09-120-653D-9	Sequence 9, Appli
43	32	28.1	18	1	US-08-351-423-3	Sequence 3, Appli
44	32	28.1	18	3	US-08-940-095-249	Sequence 249, App
45	32	28.1	18	3	US-08-940-093-249	Sequence 249, App

## ALIGNMENTS

RESULT 1  
US-08-214-770-1  
; Sequence 1, Application US/08214770  
; Patent No. 5523209  
; GENERAL INFORMATION:  
; APPLICANT: Ginsberg, Mark H.  
; TITLE OF INVENTION: METHODS FOR IDENTIFYING  
; TITLE OF INVENTION: INHIBITORS OF INTEGRIN  
; TITLE OF INVENTION: ACTIVATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,770  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 06410/002001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; US-08-214-770-1  
; Query Match 37.7%; Score 43; DB 1; Length 20;  
; Best Local Similarity 50.0%; Pred. NO. 6.8;  
; Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 KLGFPAHKKIPPEEKRE 18  
:|||||:|||||  
Db 1 KVGFFKRNRPPEEDDE 18

## RESULT 2

PCT-US95-02885-1  
Sequence 1, Application PC/TUS9502885

GENERAL INFORMATION:

APPLICANT: Ginsberg, Mark H.

APPLICANT: O'Toole, Timothy

TITLE OF INVENTION: METHODS FOR IDENTIFYING INHIBITORS

TITLE OF INVENTION: OF INTEGRIN ACTIVATION

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 55SX

OPERATING SYSTEM: MS-DOS (Version 5.0)

SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/02885

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/214,770

FILING DATE: March 14, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 06410/002001

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

PCT-US95-02885-1

Query Match 37.7%; Score 43; DB 5; Length 20;

Best Local Similarity 50.0%; Pred. No. 6.8;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 KLGFPAHKKIPPEEKRE 18  
:|||||:|||||  
Db 1 KVGFFKRNRPPEEDDE 18

## RESULT 3

PCT-US95-07542-15  
Sequence 15, Application PC/TUS9507542

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: STRUCTURAL MODELS FOR CYTOPLASMIC

TITLE OF INVENTION: DOMAINS OF TRANSMEMBRANE RECEPTORS

NUMBER OF SEQUENCES: 20

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/07542

FILING DATE: 13-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/260,514  
FILING DATE: 15-JUN-1994

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

FRAGMENT TYPE: C-terminal

ORIGINAL SOURCE:

ORGANISM: Homo sapiens integrin alpha-11b

PCT-US95-07542-15

Query Match 37.7%; Score 43; DB 5; Length 20;

Best Local Similarity 50.0%; Pred. No. 6.8;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 KLGFPAHKKIPPEEKRE 18  
:|||||:|||||  
Db 1 KVGFFKRNRPPEEDDE 18

## RESULT 4

5225193-7

Patent No. 5225193

APPLICANT: BARTPAI, TAMAS

TITLE OF INVENTION: BORDETTELLA TOXIN PEPTIDES AND VACCINES

NUMBER OF SEQUENCES: 10

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/346,837

FILING DATE: 19-OCT-1987

SEQ ID NO: 7;

LENGTH: 20

5225193-7

Query Match 34.2%; Score 39; DB 6; Length 20;

Best Local Similarity 35.0%; Pred. No. 24;

Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 1 KLGFPAHKKIPPEEKRE 20  
:|||||:|||||  
Db 1 KSEYLAHRIRIPENIRTRV 20

## RESULT 5

5225193-2

Patent No. 5225193

APPLICANT: BARTPAI, TAMAS

TITLE OF INVENTION: BORDETTELLA TOXIN PEPTIDES AND VACCINES

NUMBER OF SEQUENCES: 10

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/346,837

FILING DATE: 19-OCT-1987

SEQ ID NO: 2;

LENGTH: 20

5225193-2

Query Match 31.6%; Score 36; DB 6; Length 20;

Best Local Similarity 46.2%; Pred. No. 61;

Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 4 FPAHKKIPPEEKRE 16  
:|||||:|||||  
Db 3 YLAHRIRIPENIR 15

## RESULT 6

US-08-073-028-25  
Sequence 25, Application US/08073028

Patent No. 5464933

APPLICANT:

TITLE OF INVENTION:

TITLE OF INVENTION:

NUMBER OF SEQUENCES:

COMPUTER READABLE FORM:

MEDIUM TYPE:

COMPUTER:

OPERATING SYSTEM:

SOFTWARE:

CURRENT APPLICATION DATA:

APPLICATION NUMBER:



GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Mathews, Thomas J.  
APPLICANT: Wild, Carl T.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds,  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/073,028  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-004-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-073-028-25

Query Match 31.6%; Score 36; DB 1; Length 21;  
Best Local Similarity 43.8%; Pred. No. 65;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 7 HKKIPKREKREKLEQ 22  
Db 6 HSLIESQNGQCKNEQ 21

RESULT 7  
US-08-127-499A-15  
Sequence 15, Application US/08127499A  
Patent No. 5510264  
GENERAL INFORMATION:  
APPLICANT: VAN ALSTYNE, Diane  
APPLICANT: SHARMA, Lawrence Rajendra  
TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED  
TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/127,499A  
FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 51916/102/INBI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
US-08-127-499A-15

Query Match 31.6%; Score 36; DB 1; Length 21;  
Best Local Similarity 43.8%; Pred. No. 65;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 7 HKKIPKREKREKLEQ 22  
Db 1 HSLIESQNGQCKNEQ 16

RESULT 8  
US-08-482-847-15  
Sequence 15, Application US/08482847  
Patent No. 5556757  
GENERAL INFORMATION:  
APPLICANT: VAN ALSTYNE, Diane  
APPLICANT: SHARMA, Lawrence Rajendra  
TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR  
TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR  
TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,847  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/127,499  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 51916/104/INBI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
US-08-482-847-15

Query Match 31.6%; Score 36; DB 1; Length 21;  
Best Local Similarity 43.8%; Pred. No. 65;

	Matches	7; Conservative	3; Mismatches	6; Indels	0; Gaps	0
QY	7	HKKIPEEKREKLEQ	22			
Db	1	HSLEESQNGQEKNEQ	16			

```

RESULT 9
US-08-554-616-25
: Sequence 25, Application US/08554616
: Patent No. 6133418
:
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/554,616
FILING DATE: 06-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,028
FILING DATE: 07-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30, 742
REFERENCE/DOCKET NUMBER: 7872-004-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-554-616-25

Query Match 31.6%; Score 36; DB 4; Length 21;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0.

OY 7 HKKIPKREKREKLEQ 22
| | | | | | | | | |
Db 6 HSLIESONQOEKNEQ 21

RESULT 10
US-08-073-028-24
: Sequence 24, Application US/08073028
: Patent No. 5464933
:
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 74

```

REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-004-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-554-616-24

Query Match 31.6%; Score 36; DB 4; Length 22;  
Best Local Similarity 43.8%; Pred. No. 68;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 7 HKIPEEKREKLEQ 22  
DB 6 HSLIESONQOEKNEQ 21

RESULT 12  
PCT-US95-07542-11  
Sequence 11, Application PC/TUS9507542  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: STRUCTURAL MODELS FOR CYTOPLASMIC  
TITLE OF INVENTION: DOMAINS OF TRANSMEMBRANE RECEPTORS  
NUMBER OF SEQUENCES: 20  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07542  
FILING DATE: 13-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/260,514  
FILING DATE: 15-JUN-1994  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Synthetic peptide modeling integrin region  
PCT-US95-07542-11

Query Match 30.7%; Score 35; DB 5; Length 21;  
Best Local Similarity 44.4%; Pred. No. 88;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 1 KLGFPAAKKIPEEKREE 18  
DB 2 KVGFPKRRHATLEEDDEE 19

RESULT 13  
US-08-056-200-11  
Sequence 11, Application US/08056200  
GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul

TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESS: Knobb, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054,001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-056-200-11

Query Match 29.8%; Score 34; DB 1; Length 13;  
Best Local Similarity 66.7%; Pred. No. 76;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 12 EEEKREKEL 20  
DB 5 EEEKREKEL 13

RESULT 14  
US-07-987-272A-6  
Sequence 6, Application US/07987272A  
Patent No. 5731166  
GENERAL INFORMATION:  
APPLICANT: Geary, C., Simpson, R. J. and Lackmann, M  
TITLE OF INVENTION: No. 5731166el Chemotactic Factor  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESS: Cushman Darby & Cushman  
STREET: 1100 New York Avenue, N. W., Ninth Floor, East Tower  
CITY: Washington  
STATE: D. C.  
COUNTRY: USA  
ZIP: 20005-3918

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/987,272A  
FILING DATE: 05-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PK 2127  
FILING DATE: 05-FEB-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PK 4463  
FILING DATE: 05-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Brinkman, David W  
REGISTRATION NUMBER: 20,817  
REFERENCE/DOCKET NUMBER: DWE/1925/200259  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861 3000  
TELEFAX: 202-822 0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-987-272A-6

Query Match 29.8%; Score 34; DB 1; Length 13;  
Best Local Similarity 58.3%; Pred.No. 76;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKIPEEEKRE 17  
DB 2 AHKSHESHSKE 13

RESULT 15  
US-08-800-644-11  
Sequence 11, Application US/08800644  
Patent No. 5958752

GENERAL INFORMATION:

APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul

APPLICANT: Kim, In-Gyu

APPLICANT: Chung, Seo-Il

APPLICANT: Park, Sang-Chul

TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and

TITLE OF INVENTION: Methods of Using Same

NUMBER OF SEQUENCES: 117

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe, Martens, Olson & Bear

STREET: 620 Newport Center Drive, Sixteenth Floor

CITY: Newport Beach

STATE: CA

COUNTRY: U.S.A.

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/800,644

FILING DATE: 14-FEB-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/056,200

FILING DATE: 30-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Pedrick, Michael F.

REGISTRATION NUMBER: 36,799

REFERENCE/DOCKET NUMBER: NIH054,001A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (714) 760-0404

TELEFAX: (714) 760-9502

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 13 amino acids

TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE:  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-800-644-11

Query Match 29.8%; Score 34; DB 2; Length 13;  
Best Local Similarity 66.7%; Pred.No. 76;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 12 EEEKREKRL 20  
DB 5 EEEKREKRL 13

Search completed: July 16, 2003, 08:15:02  
Job time : 15 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 08:31:25 ; Search time 18 Seconds  
(without alignments)  
1907.586 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167

Sequence: 1 MELPVTHTLFLPLVFLTGLC.....GFFAKKIPBEKREKLEQ 1167

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size: 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	0.9	1178	1	US-08-199-776-2
2	10	0.9	1178	3	US-08-663-731-2
3	10	0.9	1178	3	US-08-879-338-2
4	10	0.9	1178	5	PCT-US95-02044-2
5	10	0.9	1178	4	US-09-293-2388-2
6	9	0.8	1065	3	US-08-630-172-9
7	9	0.8	1065	4	US-09-375-419-9
8	9	0.8	1170	2	US-08-789-078-2
9	9	0.8	1170	2	US-08-476-633-2
10	9	0.8	1170	2	US-08-476-633-2
11	9	0.8	1170	5	PCT-US95-04886-2
12	9	0.8	1170	5	PCT-US96-01314-42
13	8	0.7	17	3	US-08-893-526A-29
14	8	0.7	76	1	US-08-848-521-4
15	8	0.7	76	3	US-09-083-521-5
16	8	0.7	80	1	US-08-848-521-2
17	8	0.7	80	1	US-08-848-521-2
18	8	0.7	104	4	US-08-973-544-3
19	8	0.7	179	4	US-09-419-568F-28
20	8	0.7	179	4	US-09-354-243B-28
21	8	0.7	234	1	US-09-040-483-5
22	8	0.7	494	1	US-08-485-618-103
23	8	0.7	494	2	US-08-605-672-103
24	8	0.7	494	2	US-08-482-293A-103
25	8	0.7	494	2	US-08-943-363-103
26	8	0.7	494	2	US-09-193-043-103
27	8	0.7	494	4	US-09-688-307A-103

28	8	0.7	513	4	US-09-291-922-20	Sequence 20, Appl
29	8	0.7	1039	4	US-09-409-648-7	Sequence 7, Appl
30	8	0.7	1039	4	US-09-409-648-8	Sequence 8, Appl
31	8	0.7	1039	6	5196511-2	Patent No. 5196511
32	8	0.7	1151	1	US-08-286-889-37	Sequence 37, Appl
33	8	0.7	1151	1	US-08-485-618-37	Sequence 37, Appl
34	8	0.7	1151	1	US-08-362-652-37	Sequence 37, Appl
35	8	0.7	1151	2	US-08-605-672-37	Sequence 37, Appl
36	8	0.7	1151	2	US-08-482-293A-37	Sequence 37, Appl
37	8	0.7	1151	2	US-08-943-363-37	Sequence 37, Appl
38	8	0.7	1151	4	US-09-193-043-37	Sequence 37, Appl
39	8	0.7	1151	4	US-09-688-307A-37	Sequence 37, Appl
40	8	0.7	1152	2	US-08-476-062A-43	Sequence 43, Appl
41	8	0.7	1152	5	PCT-US96-01314-43	Sequence 43, Appl
42	8	0.7	1152	6	5424399-2	Patent No. 5424399
43	8	0.7	1153	1	US-08-173-497-3	Sequence 3, Appl
44	8	0.7	1153	1	US-08-286-889-3	Sequence 3, Appl
45	8	0.7	1153	1	US-08-485-618-3	Sequence 3, Appl
46	8	0.7	1153	1	US-08-362-652-3	Sequence 3, Appl
47	8	0.7	1153	2	US-08-605-672-3	Sequence 3, Appl
48	8	0.7	1153	2	US-08-482-293A-3	Sequence 3, Appl
49	8	0.7	1153	2	US-08-943-363-3	Sequence 3, Appl
50	8	0.7	1153	4	US-09-193-043-3	Sequence 3, Appl
51	8	0.7	1153	4	US-09-688-307A-3	Sequence 3, Appl
52	8	0.7	1155	1	US-08-286-889-46	Sequence 46, Appl
53	8	0.7	1155	1	US-08-485-618-46	Sequence 46, Appl
54	8	0.7	1155	1	US-08-362-652-46	Sequence 46, Appl
55	8	0.7	1155	2	US-08-605-672-46	Sequence 46, Appl
56	8	0.7	1155	2	US-08-482-293A-46	Sequence 46, Appl
57	8	0.7	1155	2	US-08-943-363-46	Sequence 46, Appl
58	8	0.7	1155	4	US-09-193-043-46	Sequence 46, Appl
59	8	0.7	1155	4	US-09-688-307A-46	Sequence 46, Appl
60	8	0.7	1161	1	US-08-485-618-53	Sequence 53, Appl
61	8	0.7	1161	1	US-08-485-618-55	Sequence 53, Appl
62	8	0.7	1161	1	US-08-362-652-53	Sequence 53, Appl
63	8	0.7	1161	1	US-08-362-652-55	Sequence 53, Appl
64	8	0.7	1161	2	US-08-605-672-53	Sequence 53, Appl
65	8	0.7	1161	2	US-08-605-672-55	Sequence 53, Appl
66	8	0.7	1161	2	US-08-482-293A-53	Sequence 53, Appl
67	8	0.7	1161	2	US-08-943-363-53	Sequence 53, Appl
68	8	0.7	1161	2	US-08-943-363-55	Sequence 53, Appl
69	8	0.7	1161	2	US-08-943-363-55	Sequence 53, Appl
70	8	0.7	1161	4	US-09-193-043-53	Sequence 53, Appl
71	8	0.7	1161	4	US-09-193-043-55	Sequence 53, Appl
72	8	0.7	1161	4	US-09-688-307A-53	Sequence 53, Appl
73	8	0.7	1161	4	US-09-688-307A-55	Sequence 53, Appl
74	8	0.7	1163	1	US-08-173-487-4	Sequence 4, Appl
75	8	0.7	1163	1	US-08-286-889-4	Sequence 4, Appl
76	8	0.7	1163	1	US-08-485-618-4	Sequence 4, Appl
77	8	0.7	1163	1	US-08-362-652-4	Sequence 4, Appl
78	8	0.7	1163	2	US-08-605-672-4	Sequence 4, Appl
79	8	0.7	1163	2	US-08-482-293A-4	Sequence 4, Appl
80	8	0.7	1163	2	US-08-943-363-4	Sequence 4, Appl
81	8	0.7	1163	2	US-08-943-363-4	Sequence 4, Appl
82	8	0.7	1163	4	US-09-193-043-4	Sequence 4, Appl
83	8	0.7	1163	4	US-09-688-307A-4	Sequence 4, Appl
84	8	0.7	1163	5	PCT-US96-01314-44	Sequence 44, Appl
85	8	0.7	1580	2	US-08-804-227C-11	Sequence 11, Appl
86	8	0.7	1580	2	US-08-804-198-5	Sequence 11, Appl
87	8	0.7	1891	2	US-08-804-227C-12	Sequence 12, Appl
88	8	0.7	1891	2	US-08-804-198-6	Sequence 12, Appl
89	8	0.6	9	1	US-08-462-128-4	Sequence 34, Appl
90	7	0.6	9	1	US-08-462-128-4	Sequence 34, Appl
91	7	0.6	16	1	US-08-488-212A-33	Sequence 33, Appl
92	7	0.6	16	1	US-08-320-306-33	Sequence 33, Appl
93	7	0.6	16	2	US-08-488-209B-33	Sequence 33, Appl
94	7	0.6	16	2	US-08-408-011-33	Sequence 33, Appl
95	7	0.6	17	3	US-08-893-526A-21	Sequence 21, Appl
96	7	0.6	68	1	US-07-756-250-10	Sequence 10, Appl
97	7	0.6	95	1	US-07-626-618A-2	Sequence 2, Appl
98	7	0.6	95	1	US-07-928-611-2	Sequence 2, Appl
99	7	0.6	95	1	US-08-333-977-2	Sequence 2, Appl
100	7	0.6	95	2	US-08-487-811A-2	Sequence 2, Appl

## ALIGNMENTS

```
RESULT 1
US-08-199-776-2
; Sequence 2, Application US/08199776
; Patent No. 5594120
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; TITLE OF INVENTION: No. 5594120el integrin alpha subunit
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/199,776
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-199-776-2

Query Match          0.9%; Score 10; DB 1; Length 1178;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYFSGELCP 492
|||||
DB 510 GSYFSGELCP 519

RESULT 2
US-08-663-731-2
; Sequence 2, Application US/08663731
; Patent No. 6057423
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; TITLE OF INVENTION: No. 6057423el integrin alpha subunit
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,731
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/199,776
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-663-731-2

Query Match          0.9%; Score 10; DB 3; Length 1178;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYFSGELCP 492
|||||
DB 510 GSYFSGELCP 519

RESULT 3
US-08-879-338-2
; Sequence 2, Application US/08879338A
; Patent No. 6063906
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; TITLE OF INVENTION: Antibodies to No. 6063906el Integrin Alpha
; FILE REFERENCE: B0801/7080/ERP
; CURRENT APPLICATION NUMBER: US/08/879,338A
; EARLIER FILING DATE: 1997-06-20
; EARLIER APPLICATION NUMBER: US 08/663,731
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: US 08/199,776
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1178
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (-18)...(-1)
US-08-879-338-2

Query Match          0.9%; Score 10; DB 3; Length 1178;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYFSGELCP 492
|||||
DB 510 GSYFSGELCP 519

RESULT 4
PCR-US95-02044-2
; Sequence 2, Application PC/TUS9502044
; GENERAL INFORMATION:
```

APPLICANT:  
TITLE OF INVENTION: Novel integrin alpha subunit  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield and Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02044  
FILING DATE: herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/199,776  
FILING DATE: 18 February 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1178 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-02044-2

Query Match 0.9%; Score 10; DB 5; Length 1178;  
Best Local Similarity 100.0%; Pred. No. 0.86;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYFSELCP 492  
|||||  
Db 510 GSYFSELCP 519

RESULT 5  
US-09-293-238B-2  
Sequence 2, Application US/09293238B  
Patent No. 6455042  
GENERAL INFORMATION:  
APPLICANT: Brenner, Michael B.  
TITLE OF INVENTION: A Method of Treating Ulcerative Colitis  
TITLE OF INVENTION: or Crohn's Disease by Administering an Antibody to Alpha E  
FILE REFERENCE: L0560/7005/ERP  
CURRENT APPLICATION NUMBER: US/09/293,238B  
CURRENT FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: US 08/879,338  
PRIOR FILING DATE: 1997-06-20  
PRIOR APPLICATION NUMBER: US 08/663,731  
PRIOR FILING DATE: 1996-06-14  
PRIOR APPLICATION NUMBER: US 08/199,776  
PRIOR FILING DATE: 1994-02-18  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1179  
TYPE: PRT  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: SIGNAL

LOCATION: (1)...(18)  
US-09-293-238B-2

Query Match 0.9%; Score 10; DB 4; Length 1179;  
Best Local Similarity 100.0%; Pred. No. 0.86;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYFSELCP 492  
|||||  
Db 511 GSYFSELCP 520

RESULT 6  
US-08-630-172-9  
Sequence 9, Application US/08630172  
Patent No. 6060054  
GENERAL INFORMATION:  
APPLICANT: Scaetz, Uwe  
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T  
TITLE OF INVENTION: LYMPHOCYTE VETO  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, 35th Floor  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,172  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2879-36  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1065 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-630-172-9

Query Match 0.8%; Score 9; DB 3; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574  
|||||  
Db 514 DVAVGAPLE 522

RESULT 7  
US-09-375-419-9  
Sequence 9, Application US/09375419  
Patent No. 6264950  
GENERAL INFORMATION:  
APPLICANT: Scaetz, Uwe  
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T  
TITLE OF INVENTION: LYMPHOCYTE VETO  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, 35th Floor

CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/375,419  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/630,172  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2879-36  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1065 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-375-419-9

Query Match 0.8%; Score 9; DB 4; Length 1065;  
Best local Similarity 100.0%; Pred. No. 7.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574  
|||||||  
DB 514 DVAVGAPLE 522

RESULT 8  
US-08-789-078-2  
Sequence 2, Application US/08789078  
Patent No. 5843885  
GENERAL INFORMATION:  
APPLICANT: Benedict, Stephen H.  
APPLICANT: Siahaan, Terina  
APPLICANT: Chan, Marcia  
APPLICANT: Tibbels, Scott  
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS  
STREET: 1101 Walnut St.  
CITY: Kansas City  
STATE: MO  
COUNTRY: USA  
ZIP: 64106  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/789,078  
FILING DATE: 03-FEB-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/229,513  
FILING DATE: 19-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Collins, John M.

REGISTRATION NUMBER: 26262  
REFERENCE/DOCKET NUMBER: 22833  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (816) 474-9050  
TELEFAX: 816/474-9057  
TELEX: 434-363  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1170 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: T-cell  
CELL LINE: HL-60  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..25  
OTHER INFORMATION: /label= signal  
OTHER INFORMATION: /note= "Signal sequence"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 32..79  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat I"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 82..132  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat II"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 339..391  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat III"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 392..446  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat IV"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 447..508  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat V"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 509..567  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat VI"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 568..629  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat VII"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 170..349  
OTHER INFORMATION: /label= IDomain  
OTHER INFORMATION: /note= "I-Domain"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1089..1112  
OTHER INFORMATION: /label= Trans  
OTHER INFORMATION: /note= "Transmembrane Domain"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1113..1170  
OTHER INFORMATION: /label= Cyto  
OTHER INFORMATION: /note= "Cytoplasmic domain"  
PUBLICATION INFORMATION:



AUTHORS: Pigott,  
ANTHORS: Power,  
TITLE: LFA-1 Amino acid sequence (alpha) (from  
TITLE: PMA-stimulated HL-60 cells)  
JOURNAL: The Adhesion Molecule Facts Book  
PAGES: 94-95  
DATE: 1993  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 1170  
US-08-789-078-2

Query Match 0.8%; Score 9; DB 2; Length 1170;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574  
Db 538 DVAVGAPLE 546

RESULT 9  
US-08-752-633-2  
Sequence 2, Application US/08752633  
Patent No. 5863889  
GENERAL INFORMATION:  
APPLICANT: Benedict, Stephen H.  
APPLICANT: Sahaan, Teruna  
APPLICANT: Chan, Marcia  
APPLICANT: Tibbels, Scott  
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS  
STREET: 1101 Walnut St.  
CITY: Kansas City  
STATE: MO  
COUNTRY: USA  
ZIP: 64106  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,633  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Collins, John M.  
REGISTRATION NUMBER: 26262  
REFERENCE/DOCKET NUMBER: 22833  
TELEPHONE: (816)474-9050  
TELEFAX: (816)474-9057  
TELEX: 434-363  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1170 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: T-cell  
CELL LINE: HL-60  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..25  
OTHER INFORMATION: /label= signal  
OTHER INFORMATION: /note= "Signal sequence"  
FEATURE:  
NAME/KEY: Region

LOCATION: 32..79 /label= Repeat  
OTHER INFORMATION: /note= "Repeat I"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 82..132 /label= Repeat  
OTHER INFORMATION: /note= "Repeat II"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 339..391 /label= Repeat  
OTHER INFORMATION: /note= "Repeat III"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 392..446 /label= Repeat  
OTHER INFORMATION: /note= "Repeat IV"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 447..508 /label= Repeat  
OTHER INFORMATION: /note= "Repeat V"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 509..567 /label= Repeat  
OTHER INFORMATION: /note= "Repeat VI"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 568..629 /label= Repeat  
OTHER INFORMATION: /note= "Repeat VII"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 170..349 /label= IDomain  
OTHER INFORMATION: /note= "I-Domain"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1089..1112 /label= Trans  
OTHER INFORMATION: /label= Trans  
OTHER INFORMATION: /note= "Transmembrane Domain"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1113..1170 /label= Cyto  
OTHER INFORMATION: /label= Cyto  
OTHER INFORMATION: /note= "Cytoplasmic domain"  
PUBLICATION INFORMATION:  
AUTHORS: Pigott,  
TITLE: LFA-1 Amino acid sequence (alpha) (from  
TITLE: PMA-stimulated HL-60 cells)  
JOURNAL: The Adhesion Molecule Facts Book  
PAGES: 94-95  
DATE: 1993  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 1170  
US-08-752-633-2  
Query Match 0.8%; Score 9; DB 2; Length 1170;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574  
Db 538 DVAVGAPLE 546

RESULT 10  
US-08-476-062A-42  
Sequence 42, Application US/08476062A  
Patent No. 5877275  
GENERAL INFORMATION:

APPLICANT: Arnaout, M. Amin  
TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY  
TITLE OF INVENTION: RESPONSES WITH BETA2 INTEGRINS  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,062A  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/216,081  
FILING DATE: 21-MAR-1994  
APPLICATION NUMBER: 07/637,830  
FILING DATE: 04-JAN-1991  
APPLICATION NUMBER: 07/539,842  
FILING DATE: 18-JUN-1990  
APPLICATION NUMBER: 07/212,573  
FILING DATE: 28-JUN-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Freeman, John W.  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 00786/068003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ. ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1170 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-476-062A-42

Query Match 0.8%; Score 9; DB 2; Length 1170;  
Best local similarity 100.0%; Pred. No. 7.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574  
DB 538 DVAVGAPLE 546

RESULT 11  
PCT-US95-04886-2  
Sequence 2, Application PC/US9504886  
GENERAL INFORMATION:  
APPLICANT: Benedict, Stephen H.  
APPLICANT: Stahaan, Teruna  
APPLICANT: Chan, Marcia  
APPLICANT: Tibbels, Scott  
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING  
TITLE OF INVENTION: IMMUNE TOLERANCE  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS  
STREET: 1101 Walnut St.  
CITY: Kansas City  
STATE: MO  
COUNTRY: USA  
ZIP: 64106  
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Collins, John M.  
REGISTRATION NUMBER: 26262  
REFERENCE/DOCKET NUMBER: 22833  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (816) 474-9050  
TELEFAX: 816/474-9057  
TELEX: 434-363  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1170 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: T-cell  
CELL LINE: HL-60  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..25  
OTHER INFORMATION: /label= signal  
OTHER INFORMATION: /note= "Signal sequence"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 32..79  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat I"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 82..132  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat II"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 339..391  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat III"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 392..446  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat IV"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 447..508  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat V"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 509..567  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat VI"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 568..629  
OTHER INFORMATION: /label= Repeat  
OTHER INFORMATION: /note= "Repeat VII"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 170..349  
OTHER INFORMATION: /label= IDomain  
OTHER INFORMATION: /note= "I-Domain"  
FEATURE:

NAME/KEY: Domain  
LOCATION: 1089..1112  
OTHER INFORMATION: /label= Trans  
OTHER INFORMATION: /note= "Transmembrane Domain"  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1113..1170  
OTHER INFORMATION: /label= Cyto  
OTHER INFORMATION: /note= "Cytoplasmic domain"  
PUBLICATION INFORMATION:  
AUTHORS: Pigott,  
TITLE: LFA-1 Amino acid sequence (alphaL) (from  
TITLE: PMA-stimulated HL-60 cells)  
JOURNAL: The Adhesion Molecule Facts Book  
PAGES: 94-95  
DATE: 1993  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 1170  
PCT-US95-04886-2

Query Match 0.8%; Score 9; DB 5; Length 1170;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 566 DVAVGAPLE 574  
Db 538 DVAVGAPLE 546

RESULT 12  
PCT-US96-01314-42  
Sequence 42, Application PC/TUS9601314  
GENERAL INFORMATION:  
APPLICANT: M. Amin Arnaout  
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01314  
FILING DATE: 30-JAN-96  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/380,167  
FILING DATE: 30-JAN-95  
ATTORNEY/AGENT INFORMATION:  
NAME: John W. Freeman  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 00786/267001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1170  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US96-01314-42

Query Match 0.8%; Score 9; DB 5; Length 1170;  
Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 566 DVAVGAPLE 574  
Db 538 DVAVGAPLE 546

RESULT 13  
US-08-893-526A-29  
Sequence 29, Application US/08893526A  
Patent No. 6096707  
GENERAL INFORMATION:  
APPLICANT: Heino, Jyrki  
APPLICANT: Iwasaka, Joanna  
APPLICANT: K pyl, Jarmo  
TITLE OF INVENTION: Integrin Binding Peptide and Use Thereof  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Ave., N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/893,526A  
FILING DATE: 11-JUL-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KIM, JUDITH U.  
REGISTRATION NUMBER: 40,679  
REFERENCE/DOCKET NUMBER: 1708,0240000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/371-2600  
TELEFAX: 202/371-2540  
TELEX:  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-893-526A-29

Query Match 0.7%; Score 8; DB 3; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 275 TDGESHG 282  
Db 1 TDGESHG 8

RESULT 14  
US-08-848-252-4  
Sequence 4, Application US/08848252  
Patent No. 5804177  
GENERAL INFORMATION:  
APPLICANT: Humphries, Keith R.  
TITLE OF INVENTION: METHOD OF USING CD24 AS A CELL MARKER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bereskin & Parr  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/848,252  
FILING DATE: 29-APR-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/538,052  
FILING DATE:  
APPLICATION NUMBER: US 08/151,672  
FILING DATE: 15-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McDiarmid, Shona S.  
REGISTRATION NUMBER: P-38,798  
REFERENCE/DOCKET NUMBER: 3158-028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
TELEX: 06-23115  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-848-252-4

Query Match  
Best Local Similarity 0.7%; Score 8; DB 1; Length 76;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140  
DB 12 GLLALL 19

RESULT 15  
US-09-083-521-5  
Sequence 5, Application US/09083521  
Patent No. 6048970  
GENERAL INFORMATION:  
APPLICANT: Lai, Preeti  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,521  
FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0527 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1216498  
US-09-083-521-5

Query Match  
Best Local Similarity 0.7%; Score 8; DB 3; Length 76;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140  
DB 12 GLLALL 19

RESULT 16  
US-08-848-252-2  
Sequence 2, Application US/08848252  
Patent No. 5804177  
GENERAL INFORMATION:  
APPLICANT: Humphries, Keith R.  
TITLE OF INVENTION: METHOD OF USING CD24 AS A CELL MARKER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bereskin & Parr  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/848,252  
FILING DATE: 29-APR-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/538,052  
FILING DATE:  
APPLICATION NUMBER: US 08/151,672  
FILING DATE: 15-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McDiarmid, Shona S.  
REGISTRATION NUMBER: P-38,798  
REFERENCE/DOCKET NUMBER: 3158-028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
TELEX: 06-23115  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-848-252-2

Query Match  
Best Local Similarity 0.7%; Score 8; DB 1; Length 80;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140  
DB 12 GLLALL 19

RESULT 17  
US-08-973-544-4  
Sequence 4, Application US/08973544  
Patent No. 6338950  
GENERAL INFORMATION:  
APPLICANT: WEISS, Elisabeth  
TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIKAIIDO, MARCELSTEIN, MURRAY & ORAM LLP  
STREET: 655 Fifteenth St., NW, Suite 300, G St. Lobby  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973,544  
FILING DATE: 18-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT EP 96/02663  
FILING DATE: 20-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 95109511.6  
FILING DATE: 20-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 95112201.9  
FILING DATE: 03-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Kites, Monica Chin  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P8341-7073  
TELEPHONE: (202) 638-5000  
TELEFAX: (202) 638-4810  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 97 amino acids  
TYPE: amino acid  
STRANDEDNESS: 1  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-973-544-4

Query Match 0.7%; Score 8; DB 4; Length 97;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLLA 1138  
|||||  
DB 16 LGGLLLLA 23

RESULT 18  
US-08-973-544-3  
Sequence 3, Application US/08973544  
Patent No. 6338950  
GENERAL INFORMATION:  
APPLICANT: WEISS, Elisabeth  
TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIKAIIDO, MARCELSTEIN, MURRAY & ORAM LLP  
STREET: 655 Fifteenth St., NW, Suite 300, G St. Lobby  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-5701

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973,544  
FILING DATE: 18-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT EP 96/02663  
FILING DATE: 20-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 95109511.6  
FILING DATE: 20-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 95112201.9  
FILING DATE: 03-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Kites, Monica Chin  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P8341-7073  
TELEPHONE: (202) 638-5000  
TELEFAX: (202) 638-4810  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 104 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-973-544-3

Query Match 0.7%; Score 8; DB 4; Length 104;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLLA 1138  
|||||  
DB 16 LGGLLLLA 23

RESULT 19  
US-09-419-568F-28  
Sequence 28, Application US/09419568F  
Patent No. 6331613  
GENERAL INFORMATION:  
APPLICANT: Dumoutier, Laure  
TITLE OF INVENTION: Renauld, Jean-Christophe  
TITLE OF INVENTION: (TIPS) The Proteins Encoded, and Uses Thereof  
FILE REFERENCE: LUD 5543.2  
CURRENT APPLICATION NUMBER: US/09/419,568F  
CURRENT FILING DATE: 1999-10-18  
PRIOR APPLICATION NUMBER: US09/354,243  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: US09/178,973  
PRIOR FILING DATE: 1998-10-26  
NUMBER OF SEQ ID NOS: 29  
SEQ ID NO 28  
LENGTH: 179  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
US-09-419-568F-28

Query Match 0.7%; Score 8; DB 4; Length 179;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141  
|||||

Db 21 LLLLLLV 28

RESULT 20  
US-09-354-243B-28  
Sequence 28, Application US/09354243B  
Patent No. 6359117  
GENERAL INFORMATION:  
APPLICANT: Dumoutier, Laure  
APPLICANT: Loubet, Jamila  
APPLICANT: Renaud, Jean-Christophe  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Fa  
TITLE OF INVENTION: (TIFFs)  
FILE REFERENCE: LUD-5543.1  
CURRENT APPLICATION NUMBER: US/09/354,243B  
CURRENT FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: US09/178,973  
PRIOR FILING DATE: 1998-10-26  
NUMBER OF SEQ ID NOS: 29  
SEQ ID NO 28  
LENGTH: 179  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
US-09-354-243B-28

Query Match 0.7%; Score 8; DB 4; Length 179;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLLLLV 1141  
Db 21 LLLLLLV 28

RESULT 21  
US-09-040-483-5  
Sequence 5, Application US/09040483  
Patent No. 6143867  
GENERAL INFORMATION:  
APPLICANT: Akerblom, Ingrid E.  
TITLE OF INVENTION: NOVEL HUMAN EOSINOPHIL-  
TITLE OF INVENTION: DERIVED BASIC PROTEIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,483  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/740,036  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0133 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 234 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 544241  
US-09-040-483-5

Query Match 0.7%; Score 8; DB 4; Length 234;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLLLLV 1141  
Db 4 LLLLLLV 11

RESULT 22  
US-08-485-618-103  
Sequence 103, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6500 Seaf Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-103

Query Match 0.7%; Score 8; DB 1; Length 494;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GLLLLLV 1139

Db 442 GGLLLALL 449

## RESULT 23

US-08-605-672-103  
Sequence 103, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No.5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-103

Query Match 0.7%; Score 8; DB 2; Length 494;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLALL 1139  
Db 442 GGLLLALL 449

RESULT 24  
US-08-482-293A-103  
Sequence 103, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No.5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-103

Query Match 0.7%; Score 8; DB 2; Length 494;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLALL 1139  
Db 442 GGLLLALL 449

RESULT 25  
US-08-943-363-103  
Sequence 103, Application US/08943363  
Patent No. 5837478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No.5837478el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:

CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-103

Query Match  
Best Local Similarity 100.0%; DB 2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
Db 442 GGLLLAL 449

RESULT 26  
US-09-193-043-103  
Sequence 103, Application US/09193043  
Patent No. 6251395  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6251395el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043  
CURRENT FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 103  
LENGTH: 494  
TYPE: PRT  
ORGANISM: rabbit  
US-09-193-043-103

Query Match  
Best Local Similarity 100.0%; DB 4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
Db 442 GGLLLAL 449

RESULT 27  
US-09-688-307A-103

Sequence 103, Application US/09688307A  
Patent No. 6432404  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6432404el Human Beta-2  
FILE REFERENCE: 27866/36646  
CURRENT APPLICATION NUMBER: US/09/688,307A  
CURRENT FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 08/193,043  
PRIOR FILING DATE: 1998-11-16  
PRIOR APPLICATION NUMBER: 08/605,672  
PRIOR FILING DATE: 1996-02-22  
PRIOR APPLICATION NUMBER: 08/173,497  
PRIOR FILING DATE: 1993-12-23  
PRIOR APPLICATION NUMBER: 08/286,889  
PRIOR FILING DATE: 1994-08-05  
PRIOR APPLICATION NUMBER: 08/362,652  
PRIOR FILING DATE: 1994-12-21  
PRIOR APPLICATION NUMBER: 08/943,363  
PRIOR FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 103  
LENGTH: 494  
TYPE: PRT  
ORGANISM: rabbit  
US-09-688-307A-103

Query Match  
Best Local Similarity 100.0%; DB 4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
Db 442 GGLLLAL 449

RESULT 28  
US-09-291-922-20  
Sequence 20, Application US/09291922  
Patent No. 6383776  
GENERAL INFORMATION:  
APPLICANT: Allen, Steve  
APPLICANT: Hitz, Bill  
APPLICANT: Kinney, Tony  
APPLICANT: Tingey, Scott  
TITLE OF INVENTION: Plant Sugar Transport Proteins  
FILE REFERENCE: BB-1163  
CURRENT APPLICATION NUMBER: US/09/291,922  
CURRENT FILING DATE: 1999-04-14  
EARLIER APPLICATION NUMBER: 60/083,044  
EARLIER FILING DATE: April 24, 1998  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 20  
LENGTH: 513  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-291-922-20

Query Match  
Best Local Similarity 100.0%; DB 4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1136 LLALLVFC 1143  
Db 192 LLALLVFC 199

RESULT 29  
US-09-409-648-7  
Sequence 7, Application US/09409648



Patent No. 6368797  
; GENERAL INFORMATION:  
; APPLICANT: Keith Schappert  
; TITLE OF INVENTION: METHODS FOR TREATING OR IDENTIFYING A  
; TITLE OF INVENTION: SUBJECT AT RISK FOR A NEUROLOGICAL DISEASE BY DETERMINING  
; TITLE OF INVENTION: THE PRESENCE OF A VARIANT GPIIb AND/OR VARIANT GPIIb ALLELE  
; FILE REFERENCE: 08523/015002  
; CURRENT APPLICATION NUMBER: US/09/409,648  
; CURRENT FILING DATE: 1999-10-01  
; EARLIER APPLICATION NUMBER: 60/102,624  
; EARLIER FILING DATE: 1998-10-01  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 7  
; LENGTH: 1039  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-409-648-7

Query Match 0.7%; Score 8; DB 4; Length 1039;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1130 VLGGLLLL 1137  
DB 1004 VLGGLLLL 1011

RESULT 30  
US-09-409-648-8  
; Sequence 8, Application US/09409648  
; Patent No. 6368797  
; GENERAL INFORMATION:  
; APPLICANT: Keith Schappert  
; TITLE OF INVENTION: METHODS FOR TREATING OR IDENTIFYING A  
; TITLE OF INVENTION: SUBJECT AT RISK FOR A NEUROLOGICAL DISEASE BY DETERMINING  
; TITLE OF INVENTION: THE PRESENCE OF A VARIANT GPIIb AND/OR VARIANT GPIIb ALLELE  
; FILE REFERENCE: 08523/015002  
; CURRENT APPLICATION NUMBER: US/09/409,648  
; CURRENT FILING DATE: 1999-10-01  
; EARLIER APPLICATION NUMBER: 60/102,624  
; EARLIER FILING DATE: 1998-10-01  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 8  
; LENGTH: 1039  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-409-648-8

Query Match 0.7%; Score 8; DB 4; Length 1039;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1130 VLGGLLLL 1137  
DB 1004 VLGGLLLL 1011

RESULT 31  
5196511-2  
; Patent No. 5196511  
; APPLICANT: PLOW, EDWARD F.;D'SOUZA, STANLEY E.  
; GINSBERG, MARK H.  
; TITLE OF INVENTION: PEPTIDES AND ANTIBODIES THAT INHIBIT  
; INTERLIN-LIGAND BINDING  
; NUMBER OF SEQUENCES: 31  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/444,777  
; FILING DATE: 01-DEC-1989  
; SEQ ID NO:2;  
; LENGTH: 1039  
5196511-2

Query Match 0.7%; Score 8; DB 6; Length 1039;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1130 VLGGLLLL 1137  
DB 1004 VLGGLLLL 1011

RESULT 32  
US-08-286-889-37  
; Sequence 37, Application US/08286889  
; Patent No. 5470953  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Mich  
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit.  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/286,889  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: P38,659  
; REFERENCE/DOCKET NUMBER: 27866/32168  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-286-889-37

Query Match 0.7%; Score 8; DB 1; Length 1151;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLALL 1139  
DB 1100 GGLLLALL 1107

RESULT 33  
US-08-485-618-37  
; Sequence 37, Application US/08485618  
; Patent No. 5728533  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS:  
; CITY:  
; STATE:  
; COUNTRY:  
; ZIP:  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE:  
; COMPUTER:  
; OPERATING SYSTEM:  
; SOFTWARE:  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME:  
; REGISTRATION NUMBER:  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE:  
; TELEFAX:  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH:  
; TYPE:  
; TOPOLOGY:  
; MOLECULE TYPE:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-37

Query Match 0.7%; Score 8; DB 1; Length 1151;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1132 GGLLLALL 1139  
DB 1100 GGLLLALL 1107

RESULT 34  
US-08-362-652-37  
Sequence 37, Application US/08362652  
Patent No. 5766850  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5766850e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,652  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-652-37

Query Match 0.7%; Score 8; DB 1; Length 1151;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1132 GGLLLALL 1139  
DB 1100 GGLLLALL 1107

RESULT 35  
US-08-605-672-37  
Sequence 37, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5817515e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448

TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-37

Query Match 0.7%; Score 8; DB 2; Length 1151;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
Db 1100 GGLLLAL 1107

RESULT 36  
US-08-482-293A-37  
Sequence 37, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5831029e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-37

Query Match 0.7%; Score 8; DB 2; Length 1151;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1132 GGLLLAL 1139

Db 1100 GGLLLAL 1107

RESULT 37  
US-08-943-363-37  
Sequence 37, Application US/08943363  
Patent No. 5837478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5837478e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-37

Query Match 0.7%; Score 8; DB 2; Length 1151;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
Db 1100 GGLLLAL 1107

RESULT 38  
US-09-193-043-37  
Sequence 37, Application US/09193043  
Patent No. 6251395  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6251395e1 Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043

;; CURRENT FILING DATE: 1998-11-16  
;; EARLIER APPLICATION NUMBER: 08/173,497  
;; EARLIER FILING DATE: 1993-12-23  
;; EARLIER APPLICATION NUMBER: 08/286,889  
;; EARLIER FILING DATE: 1994-08-05  
;; EARLIER APPLICATION NUMBER: 08/362,652  
;; EARLIER FILING DATE: 1994-12-21  
;; EARLIER APPLICATION NUMBER: 08/943,363  
;; EARLIER FILING DATE: 1997-10-03  
;; NUMBER OF SEQ ID NOS: 114  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 37  
;; LENGTH: 1151  
;; TYPE: PRT  
;; ORGANISM: Rattus rattus  
US-09-193-043-37  
  
Query Match 0.7%; Score 8; DB 4; Length 1151;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1132 GGLLLAL 1139  
Db 1100 GGLLLAL 1107  
  
RESULT 39  
US-09-688-307A-37  
; Sequence 37, Application US/09688307A  
; Patent No. 6432404  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, Michael W.  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 6432404el Human Beta-2  
; FILE REFERENCE: 27866/36646  
; CURRENT APPLICATION NUMBER: US/09/688,307A  
; CURRENT FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 09/193,043  
; PRIOR FILING DATE: 1998-11-16  
; PRIOR APPLICATION NUMBER: 08/605,672  
; PRIOR FILING DATE: 1996-02-22  
; PRIOR APPLICATION NUMBER: 08/173,497  
; PRIOR FILING DATE: 1993-12-23  
; PRIOR APPLICATION NUMBER: 08/286,889  
; PRIOR FILING DATE: 1994-08-05  
; PRIOR APPLICATION NUMBER: 08/362,652  
; PRIOR FILING DATE: 1994-12-21  
; PRIOR APPLICATION NUMBER: 08/943,363  
; PRIOR FILING DATE: 1997-10-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 37  
; LENGTH: 1151  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 361  
; OTHER INFORMATION: Xaa = any or unknown amino acid  
; NAME/KEY: misc\_feature  
; LOCATION: 464  
; OTHER INFORMATION: Xaa = any or unknown amino acid  
; NAME/KEY: misc\_feature  
; LOCATION: 466  
; OTHER INFORMATION: Xaa = any or unknown amino acid  
; NAME/KEY: misc\_feature  
; LOCATION: 506  
; OTHER INFORMATION: Xaa = any or unknown amino acid  
; NAME/KEY: misc\_feature  
; LOCATION: 1117  
; OTHER INFORMATION: Xaa = any or unknown amino acid  
; NAME/KEY: misc\_feature  
; LOCATION: 1118

;; OTHER INFORMATION: Xaa = any or unknown amino acid  
US-09-688-307A-37  
  
Query Match 0.7%; Score 8; DB 4; Length 1151;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1132 GGLLLAL 1139  
Db 1100 GGLLLAL 1107  
  
RESULT 40  
US-08-476-062A-43  
; Sequence 43, Application US/08476062A  
; Patent No. 5877275  
; GENERAL INFORMATION:  
; APPLICANT: Arnaout, M. Amin  
; TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY  
; TITLE OF INVENTION: RESPONSES WITH BETA2 INTEGRINS  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/476,062A  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/216,081  
; FILING DATE: 21-MAR-1994  
; APPLICATION NUMBER: 07/637,830  
; FILING DATE: 04-JAN-1991  
; APPLICATION NUMBER: 07/539,842  
; FILING DATE: 18-JUN-1990  
; APPLICATION NUMBER: 07/212,573  
; FILING DATE: 28-JUN-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Freeman, John W.  
; REGISTRATION NUMBER: 29,066  
; REFERENCE/DOCKET NUMBER: 00786/068003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1152 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-476-062A-43  
  
Query Match 0.7%; Score 8; DB 2; Length 1152;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1132 GGLLLAL 1139  
Db 1115 GGLLLAL 1122  
  
RESULT 41  
PCT-US96-01314-43

Sequence 43, Application PC/TUS9601314  
GENERAL INFORMATION:  
APPLICANT: M. Amin Arnaout  
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN  
TITLE OF INVENTION: ANTAGONISTS  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01314  
FILING DATE: 30-JAN-96  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/380,167  
FILING DATE: 30-JAN-95  
ATTORNEY/AGENT INFORMATION:  
NAME: John W. Freeman  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 00786/267001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1152  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US96-01314-43

Query Match 0.7%; Score 8; DB 5; Length 1152;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
Db 1115 GGLLLAL 1122

RESULT 42  
5424399-2  
Patent No. 5424399  
APPLICANT: ARNAOUT, M. AMIN  
TITLE OF INVENTION: HUMAN CR3a/b HETERODIMERS  
NUMBER OF SEQUENCES: 12  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/78,871  
FILING DATE: 16-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 539,842  
FILING DATE: 18-JUN-1990  
APPLICATION NUMBER: 212,573  
FILING DATE: 28-JUN-1988  
SEQ ID NO: 2;  
LENGTH: 1152  
5424399-2

Query Match 0.7%; Score 8; DB 6; Length 1152;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||

Db 1115 GGLLLAL 1122

RESULT 43  
US-08-173-497-3  
Sequence 3, Application US/08173497  
Patent No. 5437958  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No. 5437958e1 Human 2 Integrin Alpha  
TITLE OF INVENTION: Subunit  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum  
STREET: 233 S. Wacker Drive, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/173,497  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5437958and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/31363  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1153 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-173-497-3

Query Match 0.7%; Score 8; DB 1; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
Db 1116 GGLLLAL 1123

RESULT 44  
US-08-286-889-3  
Sequence 3, Application US/08286889  
Patent No. 5470953  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Mich  
TITLE OF INVENTION: No. 5470953e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum  
STREET: 233 South Wacker Drive, 6300 Seear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,889  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: P38,659  
REFERENCE/DOCKET NUMBER: 27866/32168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1153 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-286-889-3

Query Match 0.7%; Score 8; DB 1; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
Db 1116 GGLLLAL 1123

RESULT 45  
US-08-485-618-3  
Sequence 3, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1153 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-3

Query Match 0.7%; Score 8; DB 1; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
Db 1116 GGLLLAL 1123

RESULT 46  
US-08-362-652-3  
Sequence 3, Application US/08362652  
Patent No. 5766850  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,652  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1153 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-652-3

Query Match 0.7%; Score 8; DB 1; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1132 GGLLLAL 1139  
1116 GGLLLAL 1123

RESULT 47  
US-08-605-672-3  
; Sequence 3, Application US/08605672  
; Patent No. 5817515  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,672  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1153 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-605-672-3

Query Match 0.7%; Score 8; DB 2; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1132 GGLLLAL 1139  
1116 GGLLLAL 1123

RESULT 48  
US-08-482-293A-3  
; Sequence 3, Application US/08482293A  
; Patent No. 5831029  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,293A  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1153 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-482-293A-3

Query Match 0.7%; Score 8; DB 2; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1132 GGLLLAL 1139  
1116 GGLLLAL 1123

RESULT 49  
US-08-943-363-3  
; Sequence 3, Application US/08943363  
; Patent No. 5837478  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1153 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-3

Query Match 0.7%; Score 8; DB 2; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
DB 1116 GGLLLAL 1123

RESULT 50  
US-09-193-043-3  
Sequence 3, Application US/09193043  
Patent No. 6251395  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6251395el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043  
FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 3  
LENGTH: 1153  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-193-043-3

Query Match 0.7%; Score 8; DB 4; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139

DB 1116 GGLLLAL 1123

RESULT 51  
US-09-688-307A-3  
Sequence 3, Application US/09688307A  
Patent No. 6432404  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6432404el Human Beta-2  
FILE REFERENCE: 27866/36646  
CURRENT APPLICATION NUMBER: US/09/688,307A  
FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 09/193,043  
PRIOR FILING DATE: 1998-11-16  
PRIOR APPLICATION NUMBER: 08/605,672  
PRIOR FILING DATE: 1996-02-22  
PRIOR APPLICATION NUMBER: 08/173,497  
PRIOR FILING DATE: 1993-12-23  
PRIOR APPLICATION NUMBER: 08/286,889  
PRIOR FILING DATE: 1994-08-05  
PRIOR APPLICATION NUMBER: 08/362,652  
PRIOR FILING DATE: 1994-12-21  
PRIOR APPLICATION NUMBER: 08/943,363  
PRIOR FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 3  
LENGTH: 1153  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-688-307A-3

Query Match 0.7%; Score 8; DB 4; Length 1153;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
DB 1116 GGLLLAL 1123

RESULT 52  
US-08-286-889-46  
Sequence 46, Application US/08286889  
Patent No. 5470953  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Mich  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,889  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
ATTORNEY/AGENT INFORMATION:



NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: P38,659  
REFERENCE/DOCKET NUMBER: 27866/32168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-286-889-46

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 1155;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
DB 1104 GGLLLAL 1111

RESULT 53  
US-08-485-618-46  
Sequence 46, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-46

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 1155;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
DB 1104 GGLLLAL 1111

RESULT 54  
US-08-362-652-46  
Sequence 46, Application US/08362652  
Patent No. 5766850  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,652  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-652-46

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 1155;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
DB 1104 GGLLLAL 1111

RESULT 55  
US-08-605-672-46  
Sequence 46, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 5817515e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-46

Query Match 0.7%; Score 8; DB 2; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
DB 1104 GGLLLAL 1111

RESULT 56  
US-08-482-293A-46  
Sequence 46, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5831029e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-46

Query Match 0.7%; Score 8; DB 2; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
DB 1104 GGLLLAL 1111

RESULT 57  
US-08-943-363-46  
Sequence 46, Application US/08943363  
Patent No. 5837478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5837478e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-46

Query Match 0.7%; Score 8; DB 4; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
DB 1104 GGLLLAL 1111

RESULT 58  
US-09-193-043-46  
Sequence 46, Application US/09193043  
Patent No. 6251395  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6251395el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043  
CURRENT FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 46  
LENGTH: 1155  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-193-043-46

Query Match 0.7%; Score 8; DB 4; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
DB 1104 GGLLLAL 1111

RESULT 59  
US-09-688-307A-46  
Sequence 46, Application US/09688307A  
Patent No. 6432404  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6432404el Human Beta-2  
FILE REFERENCE: 27866/36646  
CURRENT APPLICATION NUMBER: US/09/688,307A  
CURRENT FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 09/193,043  
PRIOR FILING DATE: 1998-11-16

PRIOR APPLICATION NUMBER: 08/605,672  
PRIOR FILING DATE: 1996-02-22  
PRIOR APPLICATION NUMBER: 08/173,497  
PRIOR FILING DATE: 1993-12-23  
PRIOR APPLICATION NUMBER: 08/286,889  
PRIOR FILING DATE: 1994-08-05  
PRIOR APPLICATION NUMBER: 08/362,652  
PRIOR FILING DATE: 1994-12-21  
PRIOR APPLICATION NUMBER: 08/943,363  
PRIOR FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 46  
TYPE: PRT  
LENGTH: 1155  
ORGANISM: Mus musculus  
US-09-688-307A-46

Query Match 0.7%; Score 8; DB 4; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
DB 1104 GGLLLAL 1111

RESULT 60  
US-08-485-618-53  
Sequence 53, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-53

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
Db 1110 GGLLLAL 1117

RESULT 61

US-08-485-618-55  
Sequence 55, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole; Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-55

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
Db 1110 GGLLLAL 1117

RESULT 62.

US-08-362-652-53  
Sequence 53, Application US/08362652  
Patent No. 5766850  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole; Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
Db 1110 GGLLLAL 1117

US-08-362-652-53  
Sequence 53, Application US/08362652  
Patent No. 5766850  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole; Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
Db 1110 GGLLLAL 1117

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,652  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-652-55

Query Match 0.7%; Score 8; DB 1; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
Db 1110 GGLLLAL 1117

RESULT 64  
US-08-605-672-53  
Sequence 53, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-53

Query Match 0.7%; Score 8; DB 2; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
Db 1110 GGLLLAL 1117

RESULT 65  
US-08-605-672-55  
Sequence 55, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-55

Query Match 0.7%; Score 8; DB 2; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
Db 1110 GGLLLAL 1117

## RESULT 66

US-08-482-293A-53  
Sequence 53, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-53

Query Match 0.7%; Score 8; DB 2; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
Db 1110 GGLLLAL 1117

RESULT 67  
US-08-482-293A-55  
Sequence 55, Application US/08482293A  
Patent No. 5831029

GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-55

Query Match 0.7%; Score 8; DB 2; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
Db 1110 GGLLLAL 1117

RESULT 68  
US-08-943-363-53  
Sequence 53, Application US/08943363  
Patent No. 5837478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-53

Query Match 0.7%; Score 8; DB 2; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
Db 1110 GGLLLAL 1117

RESULT 69  
US-08-943-363-55  
Sequence 55, Application US/08943363  
Patent No. 5837478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-55

Query Match 0.7%; Score 8; DB 2; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
Db 1110 GGLLLAL 1117

RESULT 70  
US-09-193-043-53  
Sequence 53, Application US/09193043  
Patent No. 6251395  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6251395el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043  
EARLIER FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 53  
LENGTH: 1161  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-193-043-53

Query Match 0.7%; Score 8; DB 4; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139  
|||||  
Db 1110 GGLLLAL 1117

RESULT 71  
US-09-193-043-55  
Sequence 55, Application US/09193043  
Patent No. 6251395  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6251395el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043

```

; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-55

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 1161;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1110 GGLLLAL 1117

RESULT 72
US-09-688-307A-53
; Sequence 53, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688,307A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605,672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173,497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-688-307A-53

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 1161;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1110 GGLLLAL 1117

RESULT 73
US-09-688-307A-55
; Sequence 55, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
```

```

; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688,307A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605,672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173,497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 474
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 1128
; OTHER INFORMATION: Xaa = any or unknown amino acid
US-09-688-307A-55

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 1161;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1110 GGLLLAL 1117

RESULT 74
US-08-173-497-4
; Sequence 4, Application US/08173497
; Patent No. 5437958
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5437958el Human 2 Integrin Alpha
; TITLE OF INVENTION: Subunit
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 S. Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/173,497
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5437958and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31363
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
```



TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1163 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-173-497-4

Query Match 0.7%; Score 8; DB 1; Length 1163;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
|||||  
Db 1115 GGLLLAL 1122

RESULT 75  
US-08-286-889-4  
Sequence 4, Application US/08286889  
Patent No. 5470953  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Mich  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,889  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: P38,659  
REFERENCE/DOCKET NUMBER: 27866/32168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1163 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-286-889-4

Query Match 0.7%; Score 8; DB 1; Length 1163;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139  
|||||  
Db 1115 GGLLLAL 1122

Search completed: July 16, 2003, 08:34:39  
Job time : 23 secs

**THIS PAGE BLANK (USPTO)**